REPORT RESUMES

ED 003 879 48

THE GENERAL PHONETIC CHARACTERISTICS OF LANGUAGES.

BY- DELATTRE, PIERRE
CALIFORNIA UNIV., SANTA BARBARA
REPORT NUMBER NDEA-VI-71 PUB DATE 65
CONTRACT OEC-3-14-003
EDRS PRICE MF-$0.27 HC-$5.24 131P.

DESCRIPTORS- *SPANISH, *GERMAN, *FRENCH, *LANGUAGE, *LANGUAGE RESEARCH, LANGUAGE PATTERNS, PHONETICS, SANTA BARBARA, CALIFORNIA

THE PHONETIC FEATURES OF AMERICAN ENGLISH AND OF THE FOREIGN LANGUAGES TAUGHT IN THE UNITED STATES WERE ANALYZED AND DESCRIBED. STUDIES INVOLVED COMPARING ENGLISH TO GERMAN, SPANISH, AND FRENCH ON THE BASES OF VARIOUS PROSODIC, VOCALIC, AND CONSONANT LANGUAGE FEATURES. SPECTROGRAPHIC PATTERNS OF CONTRASTIVE UTTERANCES WERE ANALYZED, SYNTHESIZED, AND TRANSFORMED INTO SOUND BY A SPEECH SYNTHESIZER. THE EXTENT TO WHICH ASSUMED ACOUSTICAL DIFFERENCES PRODUCE APPROPRIATE AUDITORY DIFFERENCES WERE THUS DETERMINED BY LISTENING. IN ADDITION, MOTION PICTURE X-RAYS OF THE UTTERANCES WERE MADE AND STUDIED FRAME-BY-FRAME, BY MEANS OF SPECIAL PROJECTORS, TO DISCOVER THE ARTICULATORY FEATURES THAT CORRELATED WITH THE ACOUSTICAL ONES FOUND BY SPECTROGRAPHIC ANALYSIS AND SYNTHESIS. AS A COMPLEMENT TO THE INSTRUMENTAL RESEARCH, PHONETIC FEATURES OF THE FOUR SUBJECT LANGUAGES WERE INVESTIGATED BY STATISTICAL ANALYSIS, AND RELATED TO SUCH FEATURES AS PHONEME FREQUENCY, PHONEMIC DISTRIBUTION, AND SYLLABLE TYPES. RELATED MATERIAL MAY BE FOUND IN ED 003 878 AND ED 010 231. (LP)

ED003879

7/

THE GENERAL PHONETIC CHARACTERISTICS OF LANGUAGES

FINAL REPORT - 1965

Pierre Delattre

University of California
at Santa Barbara

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

RESEARCH CONTRACT: OE-5-14-003
with the U. S. Office of Education
FINAL REPORT - 1964-1965

THE GENERAL PHONETIC CHARACTERISTICS OF LANGUAGES

Pierre Delattre

Project Director

University of California
Santa Barbara

July 15, 1964 to July 15, 1965

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

TABLE OF CONTENTS

Page Number

BACKGROUND 1

OBJECTIVES 1

PROCEDURES 3

RESULTS (summarized)

A. Research published, between July 15, 1964 and July 14, 1965. 4

"Comparing the vocalic features of English, German, Spanish, and French." 4

"Comparing the consonantal features of English, German, Spanish, and French." 4

"German phonetics between English and French." 5

"Classifying speech sounds by their source." 5

B. Research completed, in press. 5

"Change as a correlate of the vowel-consonant distinction." 5

"A hierarchy of acoustic cues for the perception of speech." 6

"The acoustic synthesis of speech." 6

"Vowel nasality in French and in English." 7

"Some characteristics of German intonation for the expression of continuation and finality." 10

C. Research in progress. 13

Syllable types 13

Syllable length 13

The pharyngeal R 14

Loss of vowel color 14

Non-declarative intonation 14

TABLE OF CONTENTS
(continued)

Page Number

APPENDIX I

"Change as a correlate of the vowel-consonant distinction." 15

APPENDIX II

"A hierarchy of acoustic cues for the perception of speech." 33

Illustrations 39

APPENDIX III

"The acoustic synthesis of speech." 47

Illustrations 61

APPENDIX IV

"Vowel nasality in French and in English." 63

Illustrations 86

APPENDIX V

"Some characteristics of German intonation for the expression of continuation and finality." 69

Illustrations 116

BACKGROUND

In the teaching of foreign languages to American students one of the major problems has always been that of acquiring a satisfactory pronunciation. Language Institutes have had more difficulty in this respect than in any other. With the present emphasis on "speaking" this problem has taken on more importance every year. In order to make an effective use of the phonemic system of a second language, one must develop good articulatory habits. Improvements in the teaching of pronunciation have been hampered by an insufficient knowledge of the segmental and prosodic features of foreign languages. Problems of interference are partly due to a lack of really objective data on the phonetic features of the first language as well as of the second -- data which would allow phonetic contrasting of the two languages in a truthful and realistic manner.

OBJECTIVES

The long range objectives of this project are the instrumental analysis and detailed description of the phonetic features of American English and of the foreign languages that are commonly taught in the United States. The foreign languages toward which our main attention is turned at present are German, Spanish, and French. Results of our investigations are to appear in article or book form.

Exploratory research has lead us to divide our investigation into 40 sections -- 11 prosodic features, 13 vocalic features, and 16 consonant features. As a result, we are comparing English to German, English to Spanish, English to French, each under the 40 following headings:

Prosodic Features: 1. Declarative Intonation. 2. Non-Declarative Intonation. 3. Place of Logical Stress in the Word. 4. Place of Logical Stress in the Sense Group. 5. Nature of Logical Stress. 6. Place of Emphatic Stress. 7. Nature of Emphatic Stress. 8. Variations in Syllable Weight. 9. Internal Juncture and Syllabication. 10. Syllable Type. 11. Tension.

Vocalic Features: 12. Articulatory Description. 13. Acoustic Description. 14. New Vowel Sounds for the English Speaker. 15. Distribution (Positional and Allophonic). 16. Frequency of Occurrence. 17. Duration System. 18. Neutral Position. 19. Loss of Color. 20. Effect of Consonant Anticipation on Vowels. 21. Diphthongization. 2. Effect of Syllable Type on Vowel Color. 23. Attack and Release. 24. Nasality.

Consonantal Features: 25. Articulatory Description. 26. Acoustic Description. 27. New Consonant Sounds for the English Speaker. 28. Distribution (Positional and Allophonic). 29. Frequency of Occurrence. 30. Duration System. 31. Neutral Position. 32. Consonantal Weakening. 33. Effect of Vowel Anticipation on Consonants. 34. Speed of Articulation. 35. Tongue Fronting. 36. Aspiration. 27. Affrication. 38. Palatalization. 39. Final Release. 40. Voicing.

During the past year, our efforts were focused on the following topics, which relate either to specific portions of the long range program (4-10) or to more basic questions of speech acoustics involving the whole program (1, 2, 3):

1. The difference between vowel and consonant.
2. The hierarchy of acoustic cues in the distinctive features of speech.
3. The principle and realization of the acoustic synthesis of artificial speech.
4. The articulatory and acoustic features of vocalic nasality in French, and a comparison with nasality in English.
5. Declarative intonation in German for the expression of continuation and finality, and comparison with the same intonation in English.
6. Syllable length in relation to rhythmical patterns, place of stress, degree of stress, and syllable type (open or closed syllable).
7. A comparison of syllable types in the four languages.
8. The tongue movements of the pharyngeal R in French.
9. Loss of vowel color under the influence of weak stress, a comparison of the four languages.
10. Non-declarative intonations of German.

PROCEDURES

In order to complete these investigations with a high degree of objectivity, we have developed a three-way instrumental technique of research based on the successful design and construction of special instrumentation.

1) This three-way technique generally begins with the spectrographic analysis of utterances that have been composed and recorded for a special purpose. The comparison of spectrograms of English with corresponding ones of German, Spanish or French leads to making some hypotheses as to the acoustic differences between English and the other languages, regarding certain phonetic characteristics.

2) Then, the hypotheses are verified or refuted by means of spectrographic synthesis. Spectrographic patterns of the contrastive utterances are painted and transformed into sound by a speech synthesizer. It is thus possible to judge by ear to what extent the assumed acoustic differences produce the appropriate auditory differences.

3) Finally, motion picture X-rays of the utterances are made and studied frame by frame by means of special projectors, to discover the articulatory features that correlate with the acoustic ones found by spectrographic analysis and synthesis.

4) As a complement to this instrumental research, phonetic features of English and foreign languages are investigated by statistical analysis, related to such features as phoneme frequency, phoneme distribution, syllable types, etc.

RESULTS

Results obtained or published during the past year will be summarized here under three headings: A. Research published; B. Research completed, in press; C. Research in progress.

A. Research published between July 15, 1964 and July 14, 1965.

I. "Comparing the vocalic features of English, German, Spanish, and French." International Review of Applied Linguistics, II, 2, pp. 71-97, 15 figures. A summary of results on consonant features obtained in the first five years of investigation under O. E. contracts (1959-1964).

II. "Comparing the consonantal features of English, German, Spanish, and French." International Review of Applied Linguistics, II, 3, pp. 155-203, 30 figures. A summary of results on consonant features obtained in the first five years of investigation under O. E. contracts (1959-1964).

III. "German phonetics between English and French." Linguistics, 8, pp. 443-455, one figure. Analysis of 20 phonetic features of German indicates that in this language more features are shared with French than with English.

IV. "Classifying speech sounds by their source." In Honour of Daniel Jones, pp. 46-53, one figure. Acoustic research shows that, on the basis of their source of sound (distinguished from their resonance), speech sounds can be divided into four classes.

(Five copies of each of these four articles are attached to this final report.)

B. Research completed, in process of publication.

V. "Change as a correlate of the vowel-consonant distinction" (to appear in Studia Linguistica).

An electronic gate was built as an adjunct to our large-drum spectrograph. This gate permits one to move at will the point in time that divides an utterance into two parts and to hear the two parts separately without erasing either part. Listening experiments with this gate show that in a syllable like ba the formant transitions from the b closure to the a opening are related to the perception of the b alone, not the a; the perception of the a is only in the formant steady-states that end the syllable. These experiments confirm with natural speech what is also found with artificial speech on speech synthesizers, namely, that the perception of consonants is essentially dependent on frequency change and that of vowels on frequency stability. In order to emphasize the interest of this new theory, ten former theories on what distinguishes vowels from consonants are reviewed and their inadequacy is shown in this article. (Complete text of the article is in Appendix I.)

VI. "The hierarchy of acoustic cues for the perception of speech" (to appear in Proceedings of the Fifth International Congress of Phonetic Sciences).

The acoustic cues which compose the sounds of a language do not all possess the same distinctive power. This notion emerges from numerous experiments, either by tape splicing of natural speech or by manipulation of artificial speech -- techniques which permit one to isolate each cue and listen to its separate role in perception. We know for instance that the s friction has more distinctive power than the f friction, that the k burst has more DP than the p burst, that the k burst has more DP before rounded vowels than before spread vowels, that the second formant of velar plosives has less DP before rounded vowels than before spread ones, that the third formant of dentals has more DP than those of labials, that the closure formants of nasal consonants have less DP than their formant transitions from the viewpoint of place of articulation and more from the viewpoint of manner of articulation, etc. Eight charts of the acoustic cue DP are presented as illustrations of oppositions among manners of articulation, oppositions among places of articulation, oppositions of one consonant to all the others, and oppositions of one vowel to all the others. (Complete text of the article is in Appendix II.)

VII. "The acoustic synthesis of speech" (to appear in Bulletin de la Société des Professeurs Français en Amérique).

In order to construct a machine that will produce intelligible speech from hand-painted patterns, as a player-piano produces music from perforated bands, one must understand what sort of a musical instrument the vocal tract is and proceed by analogy. Like any musical instrument, the vocal tract has a source of sound and a resonance of that sound. The source is either periodic sound produced by the vocal cords or aperiodic sound produced by breath from the lungs forced through a narrow opening at some point between the larynx and the lips. In either case, the resonance acts as a filter which selects and passes only the sounds from the source that have the same frequency as the resonance cavities of the vocal tract. In the speech machine described here, the source is a tone-wheel that produces 50 pure tones in perfect harmonic relation (as the vocal cords would), and the selection (resonance) is made by painting the appropriate notes on a moving band of acetate. (Complete text of the article is in Appendix III.)

VIII. "Vowel nasality in French and in English" (to appear in *The French Review*).

The nasality of vowels is distinctive in French, but not in English. This contrast and the problems of interference it creates will become clear when all the linguistic aspects of nasality have been analyzed.

1. *The physiological aspect*. Motion picture X-rays of words like *sec*, *cinq*, *blâme* by five French subjects show complete agreement among the five French subjects on the following points:

a) For all nasal vowels, the velum lowers and takes the shape of an arc, whereas for nasal consonants the velum only folds laterally, keeping the shape of a right angle.

b) The four nasal vowels have the same volume of pharyngeal (back) cavity; they are only distinguished from one another by the mouth (front) cavity.

c) The tongue is clearly fronted for /ɛ̃/ and /œ̃/, but less for /œ̃/, and backed for /ɑ̃/ and /ɔ̃/, but less for /ɑ̃/.

d) The lips are clearly rounded for /œ̃/ and /ɔ̃/, but less for /œ̃/, and spread for /ɛ̃/ and /ɑ̃/, but less for /ɑ̃/.

e) The theory being that the larger the volume and the smaller the opening, the lower the note of resonance of a cavity, the front cavity resonance should be highest for /ɛ̃/, lower for /œ̃/, still lower for /ɑ̃/, and lowest for /ɔ̃/.

f) The opening of the jaws follows the same order, from widest to narrowest: /ɛ̃/, /œ̃/, /ɑ̃/, /ɔ̃/. (Note that, contrary to conventional diagrams, /ɑ̃/ is not the most open of the four nasals.)

g) From the viewpoint of the tongue, lips, and jaw positions, /ɛ̃/ and /œ̃/ differ only slightly from /ɛ/ and /œ/; but /ɑ̃/ and /ɔ̃/ differ sharply from /a/ and /ɔ/. /ɔ̃/ comes close to having the front cavity of /o/ and the back cavity of /ɔ/.

2. *The distinctive aspect*. In articulatory terms, nasal vowels are distinguished from oral vowels by the lowering of the velum and the adjusting of the pharyngeal cavity to a medium volume. Nasal vowels are mainly distinguished from one another by tongue fronting or backing and by lip rounding or spreading. In terms of cavity characteristics, distinctive features among the four nasals could be a large or a narrow front opening, and a large or a narrow back opening of the mouth cavity.

3. *The acoustic aspect*. The synthesis of nasality has demonstrated that nasal vowels are distinguished from oral ones by the low intensity of the first formant; and they can be distinguished from one another by the frequency of the second formant.

4. *The comparative aspect*. Nasality being free in English, not functional as in French, an American speaker can nasalize vowels without producing any change of meaning. Motion picture X-rays show three degrees of velum lowering in American English -- it is maximal in words like *sank*, slightly less pro-

nounced in words like Sam, and considerably less in words like sad; but in all three cases, the velic door opens near the beginning of the vowel.

5. The statistical aspect. In French, nasal vowels have a lower frequency of occurrence (3.73 percent) than oral vowels (7.09 percent). Individual frequencies are: for /ɑ̃/, 7.59 percent, for /ɔ̃/, 3.85 percent, for /ɛ̃/, 2.45 percent, and for /œ̃/, 1.05 percent. In dialects which do not use /œ̃/, the frequency of /ɛ̃/ is 3.50 percent.

6. The historical aspect. During the Gallo-Roman period, all Latin vowels nasalized when followed by a nasal consonant, even the closest ones /i. y, u/. Having nasalized they all changed color in the same direction and became mid-open with a medium back cavity. Thus [ĩ, ỹ, ũ, ã] gradually changed to /ɛ̃, œ̃, ɑ̃, ɔ̃/. It can be assumed that this change of color was brought about by an articulatory adjustment to produce optimum distinctive nasality.

7. The theoretical aspect. Nasality being produced by a lowering of the first formant intensity, the question is: how does the first formant lose intensity? X-rays show that above the pharyngeal cavity (which resonates the first formant) is a small nasal cavity called the velic cavity, the only one of the nasal cavities that has smooth enough walls to resonate (the other nasal cavities have sound-absorbant walls). When the velum is lowered and the pharyngeal cavity adjusts its volume to equal that of the velic cavity, the same harmonics can be resonated in both cavities but in opposite phase. Every pharyngeal wave that meets a velic wave of the same frequency in opposite phase is damped or cancelled.

8. The pedagogical aspect. a) Nasal vowels, having no exact correspondence in oral vowels, must be taught directly, without reference to oral vowels.

b) Nasal vowels followed by oral consonants must be pronounced without anticipation so that haut tombé will not be confused with au tome B. c) Oral vowels followed by a nasal consonant must be pronounced without anticipation so that un bonnet will not be confused with un bon nez. American habits of consonant anticipation can best be overcome by practicing minimal pairs of the b) and c) types.

9. The esthetic aspect. Poets generally call the nasal vowels obscure, veiled, nebulous. This vagueness of color can well be due to the weakness of the first formant which cannot contribute to color characterization as it does with oral vowels.

(Complete text of the article is in Appendix IV.)

IX. "Some characteristics of German intonation for the expression of continuation and finality" (to appear in Phonetica).

Intonation curves of spontaneous speech by four speakers of North German were drafted from spectrographic measurements and separated into various groups according to the attitude they expressed. The same was done for a passage read by Thomas Mann, and for a page of spontaneous speech by an American speaker for comparison with German spontaneous speech. The following results are based on those qualities which all curves for the same expression had in common. It was found that some of the contours are functional -- they distinguish finality from continuation, or minor continuation from major continuation -- while others are not functional -- they only characterize German intonation in contrast with other languages (in the present case with American English).

1. The contour of continuation begins on the last stressed syllable of a sense-group and suggests the profile outline of a bird. The tail represents the beginning of a falling pitch which reaches the bottom in the dip of the back, rises with maximum loudness in the long neck, reaches a plateau at the head and prolongs the plateau with a beak whose length depends on the number of unstressed syllables after the last stress. In short: a fall, a rise, and a high plateau.

2. The contour of finality also begins in the last stressed syllable of a sense-group. It starts like continuation, that is with a fall and a rise in the tail, back, neck, and head. But it ends with a beak pointing down if there are no stressed syllables, or with a low beak that breaks off from the head to represent the unstressed syllable, or syllables.

Functional Features

3. Finality is distinguished from continuation by the beak which points down or lies low and flat for finality, whereas it points up or remains high and level for continuation. The key to this opposition, therefore, is not heard until the unstressed portion which ends the sense-group.

4. Major continuation is distinguished from minor continuation by a deeper "back" depression, a longer "neck" rise, a higher "head," and more rise in the end of the "beak." In our recordings, the rises for minor continuation vary from a half-tone to four tones, and the rises for major continuation from one tone to eight tones.

Non-Functional Features

5. German intonation is characterized by two features.

a) A pitch depression, or dip, precedes the long rise of continuation.

b) A pitch-break separates the terminal unstressed syllables from the last stressed syllable, in finality.

Contrasting German and American Contours

6. German stressed syllables begin with a depression, (followed by a rise). English ones begin with a hump (followed by a fall) -- practically the opposite.

7. The major and loudest portion of the stressed syllable of a continuation contour rises in German but falls in English.

8. The post-stress syllables of a continuation contour are high in German but low in English.

9. The rising portion of a finality contour is substantial in German but insignificant in English.

10. The unstressed syllables after the last stress of finality break to a low level in German but are in continuity with the falling curve of the stress in English.

11. The distinction between minor continuation and major continuation is clearer, more pronounced, in German than in English.

12. The distinction between continuation and finality opposes high unstressed syllables to low unstressed syllables in German; but in English it opposes only two different degrees of fall.

Pedagogical Implications

13. American students learning German declarative intonation must acquire first the non-functional features: a) starting all stressed syllables with a depression and a rise; b) divorcing the unstressed syllables from the stressed in finality. Then they can acquire the functional features: c) distinguishing continuation from finality by maintaining the post-stress syllables high for the first, low for the second; d) distinguishing major continuation from minor continuation by marking the last stressed syllable with a deeper depression and a higher rise and the following unstressed ones with a higher plateau and more rise.

(Complete text of the article is in Appendix V.)

C. Research in progress.

X. Syllable Types

The characteristics of syllables contribute considerably to the phonic impression a language makes upon a foreign listener. We are now comparing syllable types of the four languages under eight headings:

1. Comparative statistics on the ten most frequent syllables.
2. Comparative phonetic content of the most frequent syllables.
3. Comparative ratio of consonants to vowels in syllables.
4. Comparative distribution of consonants and vowels in the syllable.
5. Comparative statistics on closed and open syllables.
6. Comparative complexity of syllable structure.
7. Comparative syllable load (number of phonemes per syllable).
8. Comparative correlation between syllable load and syllable frequency.

XI. Syllable Length

The accumulation of data on syllabic length is now being separated into two different types of results. a) On one hand it shows the correlation of syllabic length to the degree of stress (weak or strong), to the place of the syllable (final or non-final of sense group), and to the type of syllable (open or closed). b) On the other hand, it reveals the patterning of syllabic length in sense groups, and it yields a frequency of occurrence rating of the syllabic patterns of sense-groups.

For a), relationships emerge from the measurements of five minutes of spectrograms for each language (some 600 spectrograms and thousands of measurements). But for b), no clear patterns of syllable length seem to appear with the five minute material; we are attempting to make spectrograms of 30 minutes of Spanish material in the hope that the most frequent patterns of syllabic length in sense-groups will emerge.

XII. The pharyngeal R

The study of pharyngeal R by cineradiography is showing considerable variations in the place of constriction under the influence of contiguous vowels. However, this study was interrupted by an X-ray analysis of vocalic nasality in French and English for the French Review (see Appendix IV) and will soon be resumed.

XIII. Loss of Vowel Color

To compare the loss of vowel color under weak stress in the four languages, we are making spectrograms of a number of pairs of the following types, recorded by three natives of each language:

For English: compéting/ compe̲títion
For German: Fabrík/ Fabri̲kánt
For Spanish: achícan/ achi̲cában
For French: fatígue/ fati̲gánt

Frequency measurements of the first and second formants of the stressed and unstressed vowels will be the next step. Verification by synthesis will follow.

XIV. Non-declarative intonation

From a large amount of German material recorded by native actors, we have selected 743 utterances of the following types: 151 commands, 180 information questions, 185 yes-no questions, 72 implications, 39 exclamations, 92 parentheses, 24 echoes. Spectrograms of all these utterances are being made in two scales and will be analyzed to discover what all of the pitch curves expressing the same mode have in common. Similar work will be undertaken later in the three other languages for comparison. No pitch shapes are expected to emerge clearly until comparisons with pitch contours of other languages for the same mode of expression are possible.

APPENDIX I

CHANGE AS A CORRELATE OF THE VOWEL-CONSONANT DISTINCTION[1]

INTRODUCTION

We hope to show here that the new research techniques of acoustic phonetics have finally revealed the invariant contrastive cue which unambiguously distinguishes consonants from vowels. It is change versus stability. The perception of a consonant is essentially dependent on frequency change, that of a vowel on frequency stability.

The behavior of this contrastive cue can be described in two complementary statements which will be formulated now and substantiated later. A: All appreciable changes of formant frequency on a spectrographic pattern of speech, except those that connect two contiguous vowels,[2] contribute to the perception of consonants, and no appreciable change of formant frequency contributes to the perception of vowels.[3] (It is convenient to speak in terms of acoustical changes, but it is always understood that they reflect articulatory movements.) B: Whereas, in vowel/consonant contrast, vowel perception depends upon frequency stability alone (short as the acoustic segment of maximal opening may be), consonant perception may depend upon more than just frequency change -- there may be frequency stability in a consonant segment during the portion corresponding to maximal closure of the consonant constriction (not to be confused with the maxillary angle.)[4]

What can reasonably be called "changing" or "stable" was determined a few years ago at the Haskins Laboratories through a series of experiments in which [ba,ga] were transformed into [wa,ja] and finally into [u-a,i-a] by mere gradual slowing down of the rate of change of the first and second formants on speech synthesizers.[5]

The difficulty incurred in separating consonants from vowels is not phonetic; it is phonemic. As was shown by O'Connor,[6] these two classes of sounds are mutually exclusive. They normally do not contrast phonemically, as do vowels among themselves (bet/bit) or consonants among themselves (bit/pit), because their contrast necessitates a modification of the number of syllables -- French père/aère is not a mininal pair, the second element having one more syllable than the first. Yet the consonant/vowel difference may play an important role in linguistic perception as shown by Jacobson and Halle[7] who rank it as a distinctive feature.

For the convenience of this discussion, let us admit that syllables exist. They can probably best be defined as compact bundles of phonemes separated by loose contact of phonemes.[8] (Loose contact is between the last phoneme of one bundle and the first phoneme of the next bundle.) Then vowels and consonants can provisionally be assumed to exist on the basis of their distribution -- vowels are the syllable peaks, and consonants the syllable margins. Unless they are also clearly different phonetically, however, the distinction is not confirmed. Our present objective is to find acoustic and articulatory features that correlate satisfactorily with the syllabic peak vs., syllabic margin distinction.

It has often been stated that there are sounds which act sometimes like consonants, sometimes like vowels, with the implication that such sounds are phonetically alike and perceived as different only because of their distribution. It has also been said that what is distributionally a consonant in one language may be a vowel in another, again with the implication that the two sounds are phonetically the same. If these two concepts were true, the consonant/vowel contrast could not be defined by phonetic criteria but only by distributional criteria. But, according to recent acoustic research, they are not true. The two sounds which distributionally function as a consonant in one case and a vowel in the other are not phonetically the same; they have something in common, to be sure, but they are different. They are distinguished from each other by a sharp, invariant, phonetic feature: one is perceived by change in the frequency of formants, the other is not.

The crucial word in the preceding paragraphs is "perceived." Our knowledge of the phonetic content and length of a phoneme must be based on perception. Concerning length, thanks to electronic devices the frontier between two consecutive segments can be determined accurately by perception. There is no choice, therefore, but to call "consonant" the whole time-segment that is perceived as consonantal and "vowel" only the time-segment that is perceived as vocalic. Concerning the phonetic content, experiments of acoustic phonetics -- especially by synthesis -- have also thrown much light on the acoustic and articulatory correlates of vowel and consonant by perception.

A significant step forward was made when spectrographic research disclosed that the acoustic link between what corresponds to the closed and open positions of an articulation sequence, the so-called formant transitions, far from being redundant, played a key role in the perception of consonants and played no role in the perception of vowels. This makes all the kymographic and oscillographic measurements of the past obsolete because formant transitions looked like vowels on such sound recordings.

If it is acknowledged, then, that the consonant or vowel segment is acoustically limited in time by what relates to its perception, every consonant has formant transitions; every vowel has formant steady-state; and the transitions to and from the vowel steady-state entirely relate to the perception of the consonant, not to that of the vowel. The old notion (hard to remove from our minds) that a consonant is a noise, accompanied or not by voicing, was based on its closure portion only. In articulatory terms, the perception of a consonant (syllable margin) depends not only on its closure portion but also, and even more, on its articulatory movements to or from the vowel (syllable peak).

A last word of caution is necessary before the start of this discussion. To be acceptable, a criterion for the vowel/consonant contrast must apply to whispered speech as well as to loud speech. Whispered speech is perfectly intelligible and therefore contains all the acoustic cues essential to speech perception.

FORMER THEORIES

Before offering experimental evidence for the new consonant/vowel distinguishing feature of change vs., stability, let us review the major distinguishing features that are found in the linguistic literature and examine their adequacy in the light of the most recent psycho-acoustic developments. (The bibliography of former theories needs not be repeated here. It can be found in a lengthy treatment of the subject by Georges Straka: "La division du langage en voyelles et consonnes peut-elle être justifiée?"[9])

Aperture. It is said that consonants are closer (have smaller aperture) and offer more obstruction to the flow of the breath stream than vowels.

This is not true of the whole formant transition portion of consonants. It is certainly true of the closure portion, however. In the aperture during closure lies one of the most pronounced distinctions between consonant and vowel. The notorious exceptions that are often cited are only exceptions in appearance: The retroflex or bunched /r/ of American English, as in fear; the postvocalic /l/ of English, as in fill; the postvocalic /R/ of German, as in sehr; and the pharyngeal /R/ of lower class Parisian French, as in pire, all seem to "open" more than the preceding vowel. But is only the maxillary angle which widens; the constriction at the point of articulation is actually narrower for the consonant than for the preceding vowel. Motion picture X-rays of final [l] and [r] sounds show that the central tongue constriction at the alveola, palate, or pharynx, whichever is the case, is narrower for these consonants than for any adjacent vowel. They also show that the tongue is concave, centrally grooved for the vowels, whereas it is convex, centrally humped for the consonants.

In the same vein, it has been said that the /j/ of /aja/ can be more open than the /i/ of /isi/. This comparison is rather misleading, the /j/ aperture being here widened by influence of the adjacent /a/ vowels. A better criterion of aperture is the fact that in the syllables /iji/ and /uwu/ the closest portion of the consonant is always closer than the vowels.[10]

Source. It is said that vowels have their source at the glottis and resonate efficiently in the whole vocal tract, whereas consonants generally have their source above the glottis, at some point between the glottis and the lips and can resonate efficiently only in the portion of the vocal tract that is in front of the source (closure or constriction).

This is not true of the formant transition portion of consonants and is valid only for the closure portion of *voiceless* consonants. It applies less to the noise of voiced plosives and fricatives than to the noise of voiceless ones since in the former the vocal cords vibrate simultaneously with the noise and their vibrations resonate in the whole vocal tract. It does not apply measurably to nasals, glides and semi-vowels, which generally have their source at the glottis and resonate in the whole vocal tract. Besides, there can be little advantage in having the source far back; whispered vowels have their source at the glottis and their loudness is very low indeed.

Resonance. It is said that consonants do not resonate and vowels do.

This is one of the most commonly spread errors of manuals. First, the formant *transitions* of consonants resonate just like the formants of vowels. Second, non-periodic sound (friction noise, burst noise) resonates as distinctly as periodic sound (harmonics of a fundamental). It is the volume and opening of the mouth resonators which cause the perceptual distinction between the friction noises of /s/ and /ʃ/, /ç/ and /x/, /k/ and /t/, etc. If the volume of some sounds is lower than that of *voiced* vowels, it is mostly that the source of such sounds itself has a lower intensity than the source of voiced vowels. Voiceless (whispered) vowels have very low intensity at the source (narrow slit of the vocal cords), and they radiate a very low volume out of the mouth.

Periodicity. It is said that consonants have noise (non-periodic sound) and vowels do not.

This is true only of voiced and voiceless fricatives and plosives. It does not apply to nasals, glides and semi-vowels, which like voiced vowels have no noise to speak of; nor does it apply to whispered speech in which even the vowel formants are entirely non-periodic. Therefore, vowels are not distinguished from consonants by periodicity.

Formant Structure. Given that the acoustic meaning of a formant is a *narrow* concentration of energy, it is said that vowels have formant structure and consonants generally do not.

This is based on the obsolete notion that consonants are perceived by their closure-hold alone. (The closure-hold is the acoustic portion produced while the constriction is narrow, the organs being close.) Now we know that they are perceived also by their opening or closing movements. Therefore, *all* consonants have formant structure in their transitions, and many have, in addition, formant structure in their closure: nasals, glides, and semi-vowels. There are even some fricative consonants whose noise shows narrow formants similar to those of whispered vowels. The best example of such narrow formants is the [x] friction found in the German "achlaut" or the Spanish "jota."

Level of Upper Formants. It is said that when consonants have formant structure during their closure, as do nasals, glides, semi-vowels, the upper formants have lower intensity than the upper formants of vowels.

First, this does not apply to plosives and fricatives. Second, it has only relative application to the others: it is very true for nasals, less so 11 for the lateral glides, hardly at all for retroflex glides and semi-vowels.

Overall Level and Loudness. It is said that the total amplitude of a consonant, as well as its resulting loudness, is lower than that of a vowel.

This is mostly true when applied to the closure-hold or the noise of the consonant but much less so when applied to its formant transitions which reach an intensity equal to that of the adjacent vowel well before the steady-state is reached. Even during closure, the intensity and loudness of a retroflex or bunched /r/ are equal to those of many vowels. It is precisely because of the high intensity of their formant transitions that many consonants are identified by their transitions rather than by their noise, (explosion or friction) since noise is too low in loudness and power of transmission.

Air Pressure. It is said that even though the loudness of a consonant is generally lower than that of a vowel, the air pressure from the lungs is usually greater for the consonant than for the vowel.

Naturally this applies to the closure phase, not to the formant transitions, and is maximal in voiceless fricatives and plosives, minimal in nasals and glides.

Articulatory Energy. It is said that there is a greater expense of articulatory energy for the consonant than for the vowel.

This is especially true when the consonant is a voiceless plosive or fricative. However, it does not clearly apply to the most periodic consonants, such as nasals and glides, and it leaves out the case of whispered vowels for which more energy is expended than for any consonant.

Yield. It is said that consonants have a lower yield than vowels; that for an equal articulatory effort they produce less sound volume; that in producing a consonant the vocal tract is a less efficient transformer of articulatory energy than in producing vowels.

This does not apply to whispered speech where consonants even have a higher yield than vowels. The main reason for high return in speech sounds is the great efficiency of the vocal cords when they vibrate so that all voiced sounds, including voiced consonants, are more efficient than their voiceless counterparts. Whispered vowels have the lowest yield of any speech sounds. Whispered /a/ is a good rival of /θ/ when it comes to inefficiency or low yield!

Direction of Articulatory Movements. The most recent contribution to explaining the consonant/vowel dichotomy was offered by Georges Straka in the opening volume of *Travaux de Linguistique et de Littérature de l'Université de Strasbourg* quoted earlier. Having hypothesized that opening muscles operate for vowels and closing muscles for consonants (on the dubious basis of muscular behaviour in mastication) and having observed that in emphatic articulation vowels are more open and consonants more close than in normal articulation, Straka sees vowels as "ouvreuses" and consonants as "fermeuses."

Interesting as this difference is, because of its categorical character, it disagrees with what perception teaches us since it implies that vowels, as well as consonants, are perceived by articulatory movements (change). The fact is that consonants are "ouvreuses" in syllable-initial position when their formant transitions are "releasing," and "fermeuses" in syllabic-final position when their formant transitions are "arresting," and that vowels are neither "ouvreuses" nor "fermeuses" since no articulatory movement contributes to their perception.

EXPERIMENTAL EVIDENCE

Experimental evidence of the fact that consonants are perceived by change and vowels by stability abounds today. But it is all recent. It is found entirely in the spectrographic research of the last decade or so. Let us summarize here some major types of evidence.

1. Research on the role of formant transitions for the perception of English consonants has been reported in a score of papers during the past decade and

a half.[12] Those articles show that in English, whereas all consonants need formant transitions to be clearly identified, only six, /s/, /z/, and /ʃ/, /ʒ/ before all vowels, and /k/, /g/ before rounded vowels, really require noise in addition to transitions:[13] for instance the friction of /ðo/ and the burst of /bo/, have extremely low loudness and seem to play no significant part in their identification. Their manner of articulation is largely perceived by the direction of first formant transitions and the tempo of all formant transitions; their place of articulation is largely perceived by the direction of second and third formant transitions as defined by their respective loci.[14]

2. When a spectrogram, or spectrographic pattern, is being moved under the lights of a speech synthesizer of the dynamic type, the portions that show frequency changes in the formants contribute to the identification of consonants; those that show relative steady state in the formant frequency, even for a very short time, are identified as vowels.

3. When the speech synthesizer is stopped at any point along the formant transitions of a consonant pattern, the sound produced is that of a vowel. The reason is, of course, that during the stop there is no frequency change. (Unlike a magnetic tape or a phonograph record, a speech synthesizer can stop its tape without interruption of sound or can slow down to any speed without change of fundamental pitch.) For instance, if the synthesizer is stopped at intervals of 0.02 seconds during the conclusion of the final /l/ of fall, one hears a succession of vowels ranging from an approximate [ɔ] to an approximate [U].

4. As the speed at which a given speech pattern passes under the lights of a synthesizer is appropriately changed, some phonemes change from one class to another. Thus a given slowing down of /ga/, /ba/, changes them to /ja/, /wa/, and a further slowing down changes them into disyllabic /i-a/ /u-a/. This

happens because the rate of change of formant transitions (their tempo) is appropriate to a different class of phonemes at each different speed. When /ga/, /ba/, is heard the formant transitions have a tempo of plosives; when it is /ja/, /wa/, they have a tempo of semi-vowels; when it is /i-a/, /u-a/, the formants have a tempo of vowels in which change is negligible.

5. We have built, as an accessory to our spectrograph, an electronic time-gate which permits one to move at will the point in time that divides an utterance into two parts and to hear the two parts separately without erasing either part. For example, we can divide /ba/ into two parts and move the dividing point back and forth until the first part is heard as [b] and the second as [a]. When this is done, for every American sequence of "consonant + vowel" a satisfactory dividing point is found with the consonant as well as the vowel maximally identifiable. This dividing point is the farthest one to the right at which the consonant can be heard without the least suggestion of the following vowel. On the spectrogram this dividing point puts the transitions on the side of the consonant, and the vowel is in near steady-state. As the dividing point is moved to the left, the intelligibility of the consonant deteriorates; when the transition has finally been eliminated from the left portion, the consonant can no longer be identified.

6. Using this device to locate the phonemic frontiers in words like bird /bɝd/ and rare /rɛr/, we find that the syllabic peak of /ɝ/ in bird is not perceived by transitions -- the transitions are all in the /b/ and the /d/ -- whereas the syllabic margins of /r/ in rare are perceived by pronounced transitions. The same sort of experiment can serve to compare yet with eat. Consonantal /j/ is perceived by transitions; vocalic /i/ is not.

This last experiment brings us back to our introduction: central /ɝ/ and marginal /r/ are not phonetically the same as is often implied is manuals.

Marginal /r/ is perceived by change in its formant frequencies; central /ɚ/ is not. In other words, the segment /r/ includes formant transitions; the segment /ɚ/ does not.

MARGINAL CASES

1. When two vowels, both syllable peaks, occur successively without interruption as in *May eats* [me-its], the two vowels are connected on the spectrogram by uninterrupted formants which change their frequency from those of [e] to those of [i] and therefore can legitimately be called formant "transitions" but which do not relate to the perception of a consonant. This observation compels us to distinguish *neutral* formant transitions from *consonantal* formant transitions. In *May eats* the transitions preceeding [i] are neutral; in *May beats*, they are consonantal. (There are no vocalic transitions since vowels are perceived by steady-state.) The best sign by which to recognize one type from the other on spectrograms is found in the first formant, which is the one mainly responsible for the perception of the *manner* of articulation. In neutral transitions the first formant moves directly from vowel to vowel. In consonantal transitions the first formant moves to and/or from a closure-position. In *May beats*, for instance, the first formant will fall sharply, then rise sharply. But the opposite may also occur. In the French word *mairie* [meri], because the /r/ is articulated in the back half of the vocal tract the first formant will *rise* from the [e] position, and then fall to the [i] position.

This is perhaps the place to mention that in distinguishing classes of sounds, the first formant plays the major role. For instance, it is the differences in *tempo* of first formant which distinguish most effectively [ga], [ja] and [i-a] from each other.

2. The case of formant transitions which end falling diphthongs is an ambiguous one which may need an arbitrary decision. On the basis of what has

been said, these diphthongal transitions might be classed as consonantal because they show a sign of closure in their first formant, although very weakly at times. For instance, the falling diphthong [ai] is very different, on spectrograms, from the sequence [a-ɪ] in two syllables. The former ends in an embryonic fall in the first formant; the latter does not. Another reason for considering the diphthongal falling glide as consonantal is that when the diphthong is played in reverse, its falling part is generally perceived as a consonant. When recordings of *my*, *say*, are played in reverse, they are generally heard as [jam], [jɛs].

3. Clusters of three consonants may appear to create another problem. One may object that the /ʃ/ of *reached* [ritʃt] and the /k/ of *risks* [rɪsks] have no contiguous vowels from which or to which formant transitions can operate, yet they are distinct from [ritst] or [rɪsps]. Spectrographic analysis, as well as synthesis, provides a clear answer to this objection. a) Friction noises have significant transitions in the noise itself. Their lower limit is seldom horizontal on spectrograms. In synthesis, the sequence [tʃt] can be modified to be heard as [tʃp] by merely changing the angle (falling instead of rising) of the noise's lower limit. Similarly, [sks] can be changed into [sps] by modifying the angle of the noise's lower limit. b) It is doubtful that sequences of three consonants actually occur without an intrusive vowel being formed either between the first and the second consonant or between the second and the third. This vowel may be whispered or voiced. c) In synthesis, using a neutral friction noise, [tʃt] can be changed to [tst] merely by changing the locus of the third formant transition before and after the [ʃ]. d) It is also possible that transitions do not have to go exclusively from vowel to consonant or consonant to vowel. It was shown in part 1 that transitions can be homovocalic. They perhaps can also be homoconsonantal -- a consonant may function as a vowel for a contiguous consonant. For instance, in [tʃt], the [t] bursts may well be using the [ʃ] friction in place of a vocalic partner, for the [t] bursts alone are not understandable. e) Finally, the role

of context in the identification of clusters of three consonants is very important, minimal pairs with 3 consonants being very scare. Redundancy is therefore high. Since [pɝtst] and [ritst] do not exist in English, in fluent speech they would be understood as perched and reached.

4. It cannot be denied that the friction sounds of [ʃʃʃ] and [sss] can be distinguished from each other when produced by mouth. But are those noises "consonants"?

a) We know that the noises of fricatives are much less intelligible when the initial and final portions -- which include some sort of transition -- are severed from the tape recording of those sounds. b) In actual speech there is always more than the mere noise involved in consonantal perception. c) It is also known that, using the same noise but varying the transitions, one can easily transform [ʃ] to [s], in front of any vowel.[16]

5. Many spectrographic patterns of syllables seem to show no steady-state. Yet it is certain that the ear catches one instant of at least 3 hundredths of a second -- the minimum time necessary to identify vowel color -- as the syllable peak. Experiments with a time-gate indicate that this instant is located between reversals of direction in the first formant. An example of such extreme cases is the syllable [bɛg]. Because of the very low locus of the /b/ and the very high locus of the /g/, the second formant shows a continuous rise from beginning to end with no stop visible at the syllable peak. The first formant, however, rises, flattens, and falls; and its flat top is located precisely where the vowel color /ɛ/ can be identified with an electronic gate. Moreover, the third formant describes a rising-falling arc, the top of which coincides with the top of the first formant.

The other most extreme case known to spectrographers is the syllable /bɛd/. There it is the third formant which rises sharply throughout because of the low locus of /b/ and the high locus of /d/. But both the first and second formants rise and fall, and the syllable peak which identifies the vowel /ɛ/ is easy to locate at the hump of both arcs. In short, experiments with a time-gate indicate that when on spectrograms of a syllable, formants do not show even an instant during which all frequencies are stable, the ear identifies the vowel color of the syllable peak by the segment located between reversals of direction in the first formant.

NUMBER OF CLASSES

The fact that consonants have formant transitions and vowels have not, does not mean that the division into two classes is the only possible one. The /ga-ja/ experiment shows that two or more consonant classes may be distinguished by tempo of the transitions, and it so happens that the slow-tempo consonants are distributionally different from the fast-tempo consonants; the slow-tempo consonants can only occur adjacent to a vowel. We can then have at least three classes of sounds in a syllable: a) the peaks (vowels), b) the margins that can only be located adjacent to a peak (semi-vowels and the closing portions of diphthongs), and c) the margins that can be located one step or more from the peak (the other consonants). Their respective phonetic correlates are: no change, slow tempo change, and fast tempo change. And we could perhaps add one more class in English distinguishing the pre-peak margins adjacent to a peak, with slow tempo as /w/ in swam /swam/, from the post-peak margins adjacent to a peak, with extra-slow tempo, as /w/ in mouse /maws/.

Footnotes

CHANGE AS A CORRELATE OF THE VOWEL-CONSONANT DISTINCTION

[1] The research reported herein was performed pursuant to a contract with the U. S. Office of Education, Department of Health, Education, and Welfare.

[2] The formant transitions which connect one vowel to another are the only truly redundant ones -- the only ones that actually should bear the name of "transition". They differ from the significant formant transitions which connect a vowel aperture to a consonant closure or a consonant closure to a vowel aperture in that they show no sign of closure-like lowering of the first formant.

[3] If, in extreme cases, a short vowel can be perceived between two consonants along a continuously changing pattern, as for instance in [bɛg], it is not because of the frequency changes that precede and follow the point at which the vowel is perceived but in spite of them. The instant (between the preceding and the following consonant transitions) at which the vowel is perceived is designated by the occurrence of a reversal in the first formant direction. In other words, the point at which the vowel is perceived stands between a rise and a fall, or a fall and a rise, of the first formant. This instant is short enough to function as a vowel segment of stable frequency and long enough to make possible the identification of vowel color (a minimum of 3 centiseconds).

[4] For instance, during the closure of consonants like [l] or [m], the formant frequencies often are nearly stable. Furthermore, making the frequency of all formants completely stable in synthesized patterns does not interfere with the perception of the complete segment (stable portion plus formant transistions) as consonantal.

Footnotes

CHANGE AS A CORRELATE OF THE VOWEL-CONSONANT DISTINCTION

[5] A. M. Liberman, P. Delattre, L. Gerstman and F. S. Cooper, "Tempo of frequency change as a cue for distinguishing classes of speech sounds," *J. Exp. Psychol.*, 52, pp. 127-138, 1956.

[6] J. D. O'Connor and J. L. M. Trim, "Vowel, consonant, and syllable -- a phonological definition," *Word*, 9, 2, pp. 103-122, August, 1953.

[7] R. Jakobson, G. M. Fant and M. Halle, *Preliminaries to Speech Analysis*, Boston, M. I. T., 1952.

[8] B. Malmberg, "Remarks on a recent contribution to the problem of the syllable," *Studia Linguistica*, 15, 1, pp. 1-9, 1961.

[9] *Travaux de Linguistique et de Littérature*, published by Le Centre de Philologie et de Littératures Romanes de l'Université de Strasbourg, pp. 17-99, 1963.

[10] Good spectrograms of such syllables can be found in "Acoustic characteristics of selected English consonants," by Ilse Lehiste, *IJAL*, 30, 3, p. 125, 1964. In [ji] the first formant is lower and the third higher for [j] than for [i], an indication that the constriction between tongue and palate is narrower for [j] than for [i]. In [wu], the first and second formants are lower for [w] than for [u], an indication that the constriction at the lips is more pronounced for [w] than for [u].

[11] The last two contrastive cues are mentioned by M. Halle in the *Sound Pattern of Russian*, The Hague, Mouton, 1959.

Footnotes

CHANGE AS A CORRELATE OF THE VOWEL-CONSONANT DISTINCTION

[12] A. M. Liberman, P. Delattre, F. S. Cooper and L. Gerstman, "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," Psychol. Monogr., 379, pp. 1-14, 1954.

P. Delattre, A. M. Liberman, F. S. Cooper, "Acoustic loci and transitional cues for consonants, " JAS., 27, pp. 769-774, 1955.

B. Malmberg, "The phonetic basis for syllable division," Studia Linguistica, 9, pp. 80-87, 1955.

K. Stevens and A. House, "Studies of formant transitions using a vocal-tract analog," JAS., 28, pp. 578-585, 1956.

J. D. O'Connor, L. Gerstman, A. M. Liberman, P. Delattre and F. S. Cooper, "Acoustic cues for the perception of initial /wjrl/ in English, "Word, 13, pp. 24-44, 1957.

P. Delattre, "Les indices acoustiques de la parole: Premier rapport," Phonetica, 2, 1/2, pp. 108-118: 2, 3/4, pp.226-251, 1958.

P. Delattre, "Le jeu des transitions de formants et la perception des consonnes," Proceedings of the Fourth International Congress of Phonetic Sciences, The Hague, Mouton, pp. 407-417, 1962.

P. Delattre, A. M. Liberman and F. S. Cooper, "Formant transitions and loci as acoustic correlates of place of articulation in American fricatives," Studia Linguistica, 16, 2, pp. 104-121, 1964.

K. S. Harris, H. S. Hoffman, A. M. Liberman, P. Delattre, and F. S. Cooper, "Effect of third-formant transitions on the perception of the voiced stop consonants," Jour. Acoustical Soc. Amer., 30, pp. 122-126, February, 1958.

[13] Katherine S. Harris, "Cues for the discrimination of American English fricatives in spoken syllables, " Language and Speech, 1, pp. 1-7, 1958 .

Footnotes

CHANGE AS A CORRELATE OF THE VOWEL-CONSONANT DISTINCTION

[14] P. Delattre, A. M. Liberman and F. S. Cooper, "Acoustic loci and transitional cues for consonants," JAS., 27, pp. 769-774, 1955.

[15] On the role of first formant as a cue for place of articulation, see p. 415 of "Le jeu des transitions de formants," Proceedings of the Fourth International Congress of Phonetic Sciences, The Hague, Mouton, 1962.

[16] This was done repeatedly for the experiments described in "Formant transitions and loci as acoustic correlates of place of articulation in American fricatives," Studia Linguistica, 16, 2, pp. 104-121, 1964.

APPENDIX II

A HIERARCHY OF ACOUSTIC CUES FOR THE PERCEPTION OF SPEECH

(translated from the French)

The knowledge of acoustic cues which make up the distinctive elements of a language is now well enough advanced to allow taking stock of our position on the general problem of the distinctive power of each of these cues. The relativity of this power has been observed indirectly or partially in a number of studies carried on by varied techniques. What remains for us to see here is how to proceed to a systematic investigation of this hierarchy.

The method which proceeds by cutting and splicing magnetic tapes on which human speech has been recorded is the one used by André Malécot to study the resonances of nasal consonants during the buccal closure (that we shall call here CLOSED RESONANCES) and determine their role in distinguishing the place of articulation. [1] By this same method, Katherine Harris compared the different noises of fricative consonants to each other in order to study their contribution to the identification of the place of articulation. [2] Lee Lisker varied the duration of the buccal closure to discover its role in the voicing of stop consonants; [3] Frederick Householder, [4] Morris Halle, [5] and others have compared the perception of final stops with and without their explosive release. But the method of tape splicing is limited in two ways. First it allows only for reduction and addition, it does not allow for the modification of an acoustic cue. Next it allows the cues to be broken down into segments which have only a sequential, but not a parallel, relation to time. In order to leave out, change, manipulate and even isolate the acoustic cues of speech, maintaining a parallel relation to time, one must proceed by synthesis of artificial speech. Thus, it is above all by means of synthetic speech that the contribution of each cue to the perception of place and manner of articulation can be shown. Here the most carefully worked-out research technique consists in varying one single cue in time, in frequency and in intensity, while the other cues are "neutralized." Thus the researchers of Haskins Laboratories were able to study, for English, the burst noises of oral stop consonants,[6] the transitions of nasal and stop consonants,[7] the closed resonances and the transitions of resonants,[8] the transitions and the noises of fricatives,[9] vowel formants,[10] the effects of tempo,[11] etc.

As a result of all this impressive work, many facts about hierarchy are known to us. We know for example that, for the perception of the place of articulation en English, noise plays a greater part in the perception of voiceless consonants than of voiced ones, and that on the other hand the transitions play a more important role in the voiced consonants than in the voiceless ones. But these noises have very different distinctive values: friction noises are more efficient than burst noises; among the frictions, the most useful are the most intense, those of /ʃ/, /s/; when the frictions are weak, as for /θ/, /f/, the place of articulation is revealed by the transitions more than by anything else. Among the burst noises, the ones that are the most useful to perception are those of velar consonants followed by rounded vowels -- thus in /go/, it is principally the noise that identifies the place of articulation as velar, in /bo/ it is principally the second formant. We know that the third transition plays a more important role in the retroflex r̲ or the bunched r̲ than in other resonants, and is more critical in dentals than in labial or velar consonants. We know that the place of articulation of the nasal consonants is identified by the transitions much more than by the closure formants. We know that the tempo (rate of change) of the transitions helps to distinguish /y/ from /g/, /ð/ from /v/, but not /z/ from /ʒ/. And so on. Isolated indications of the relative power of acoustic cues are abundant; our present concern is to arrange them in some order.

To achieve an over-all view of this hierarchy we have set up charts on which known cues come into play to distinguish between the two terms of all the oppositions of phonemes, of places of articulation, of manner of articulation and of voicing in a given language. These charts (see figures) are based on four different kinds of oppositions. We do not have enough space to give more than one or two samples of each type of opposition. The distinctive power of a given cue for a given opposition is noted with X marks in the appropriate square. The notation has been made arbitrarily in five degrees, from no X to four X's. The judgements as to the amount of distinctive power are based largely on published works, but also on exploratory research by the author. The examples of consonants are taken from English, except that the affricated consonants have been omitted because of the ambiguous unity they present. Therefore, there remain four manners of articulation: the plosives, the fricatives, the nasals and the resonants, this last manner including liquids and semi-vowels. The examples of vowels are taken from French so as to include the cues for nasality.

The first type of opposition involves opposing successively each MANNER of articulation to all the other manners (Figure 1). For four manners this makes six oppositions. For example, on the middle horizontal line one can see that the tempo of the transitions is more useful to the stop/resonant distinction than to other distinctions (stop consonants = plosives and nasals); and in the first vertical column one can see that plosives are distinguished from nasals especially by the closed resonance and the first transition.

Our second approach is to oppose the PLACES of articulation to each other separately, within each manner (Figure 2). It is necessary to consider velar consonants before spread vowels (ge, ŋe) separately from velar consonants before rounded vowels (ga, ŋa), since the latter do not make use of a second formant velar locus for the perception of the velar place of articulation. Several experiments prove the above. The simplest of them is performed by cutting out the burst noise from a recording of /go/. One then hears /bo/, because the second transition of /go/ goes toward the labial locus. This also demonstrates that the burst noise of /go/ plays a major role in the perception of the velar place of articulation, as we see noted on Figure 3, line 3, sounds /ka/ and /ga/.

Our third approach is to set up a single consonant, /m/, in opposition to s all the other consonants (Figure 3). This brings to light the numerous cues which come into play in English: 6 for place of articulation (for French, German and Spanish a 7th must be added, that of the locus of the first transition, since these languages have pharyngeal consonants at a very high locus), 5 for manner of articulation, and 7 for voicing (according to current study).

Also in the third approach we set up a vowel, the French nasal /ɛ/, in opposition to all the other French vowels (Figure 4). The first formant stands out in this opposition as very important because it plays a double role, both through its frequency variations in the determination of aperture and through its intensity variations in the perception of nasality.

Our fourth approach is to oppose all the consonants to each other for a single acoustic cue at a time. Here we have chosen the cue of NOISE (Figure 5) and the cue of the SECOND TRANSITION (Figure 6). The comparison of these two charts leaves no doubt about the fact that the distinctive power of the second transition is more extensive. The distinctive power of noise is great only for four fricatives: /ʃ/, /s/, /ʒ/, /z/, and two plosives: /g/ and /k/ before rounded vowels.

Again in our fourth approach, we oppose all the French vowels to each other by a single acoustic cue at a time. We have chosen for purpose of illustration the frequency of the first formant (Figure 7) and of the second formant (Figure 8). In comparing these two charts, we see that the power of the second formant (frequency) is slightly more extensive than that of the first formant (frequency), both because of the nasal vowels which are only distinguished from each other by the frequency of the second formant, and also because of the fact that the second formants coincide less frequently than do the first formants.

Let us conclude. The problem of the hierarchy of acoustic cues is vast. We have only been able to give an over-all view of it by showing what kind of charts must be set up so that we can see how to proceed in this study. These charts are based on four different kinds of oppositions. To complete the present outline on the third level of opposition , we would need as many charts as there are consonants and vowels in the language to be analyzed. For the fourth kind of opposition, there would have to be as many charts as there are acoustic cues that can be isolated. Moreover, we have only examined here the oppositions of segmental phonemes. A comprehensive study would also include an examination of the acoustic cues of the prosodic phonemes of accent, intonation and juncture, as well as those of the consonant/vowel dichotomy, whether phonemic or not.

Eventually it will be interesting, once the charts are complete, to group the cues by family -- for example, to compare the frequency cues with those of intensity, of duration and of tempo; or even to compare them according to their articulatory correlates.

Bibliography

[1] A. Malécot: Acoustic Cues for Nasal Consonants; an Experimental Study Involving a Tape-Splicing Technique. Language 32:274-284 (1956).

[2] Katherine Harris: Cues for the Discrimination of American English Fricatives in Spoken Syllables. Language and Speech 1:1-7 (January 1958).

[3] L. Lisker: Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English. Language 33:42-49 (1957).

[4] F. Householder: Unreleased ptk in American English. For R. Jakobson, pp. 235-244 (1956).

[5] M. Halle, G. Hughes, J.-P. Radley: Acoustic Properties of Stop Consonants. Journ. Acoust. Soc. America 29:107-116 (1957).

Bibliography (continued)

P. Delattre: Unreleased Velar Plosives after Back-Rounded Vowels. JAS 30, 6:581-582 (June 1958).

6 A. Liberman, P. Delattre, F.S. Cooper: The Role of Selected Stimulus-Variables in the Perception of the Unvoiced Stop Consonants. Amer. J. Psychol. 65:497-517 (1952).

7 P. Delattre, A. M. Liberman, F. S. Cooper: Acoustic Loci and Transitional Cues for Consonants. JAS 27:769-774 (1955). 5

K. Harris, H. Hoffman, A. Liberman, P. Delattre, F. Cooper: Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants. JAS 30, 2:122-126 (February 1958).

Howard S. Hoffman: Study of Some Cues in the Perception of Voiced Stop Consonants. JAS 30, 11:1035-1041 (November 1958).

A. Liberman, P. Delattre, F. Cooper, L. Gerstman: The Role of Consonant-Vowel Transitions in the Perception of the Stop and Nasal Consonants. Psychol. Monogr. 379:1-14 (1954).

8 J. O'Connor, L. Gerstman, A. Liberman, P. Delattre, F. Cooper: Acoustic Cues for the Perception of Initial /wjrl/ in English. Word 13:24-44 (1957).

9 P. Delattre, A. Liberman, F. Cooper: Formant Transitions and Loci as Acoustic Correlates of Place of Articulation in American Fricatives. Studia Linguistica (1963).

10 P. Delattre, A. Liberman, F. Cooper, L. Gerstman: An Experimental Study of the Acoustic Determinants of Vowel Color; Observations on One- and Two-Formant Vowels Synthesized from Spectrographic Patterns. Word 8:195-211 (1952).

11 A. Liberman, P. Delattre, L. Gerstman, F. Cooper: Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds. J. Exp. Psychol. 52:127-138 (1956).

	Plosives / Nasals	Plosives / Fricatives	Plosives / Resonants	Nasals / Fricatives	Nasals / Resonants	Fricatives / Resonants
Continuity / Discontinuity		$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	
Locus of First Transition	$\overset{xx}{xx}$	$\overset{x}{x}^{x}$	$\overset{x}{x}^{x}$	$\overset{x}{x}^{x}$	$\overset{x}{x}^{x}$	x_x
Transition Tempo		x_x	$\overset{x}{x}^{x}$	x_x	$\overset{x}{x}^{x}$	x
Closed Resonances	$\overset{xx}{xx}$	x_x	$\overset{xx}{xx}$	x_x	$\overset{xx}{xx}$	x_x
Noise Duration	x	$\overset{x}{x}^{x}$	x	$\overset{x}{x}^{x}$		$\overset{x}{x}^{x}$

FIGURE 1

Distinctive power of the acoustic cues of MANNER OF ARTICULATION in English. Comparison of the four manners to each other, forming six pairs of oppositions.

	b/d	b/ge	b/ga	d/ge	d/ga	m/n	m/ŋe	m/ŋa	n/ŋe	n/ŋa	v/ð	v/z	v/ʒ	ð/z	ð/ʒ	z/ʒ	w/j	w/r	w/l	j/r	j/l	r/l
Locus of Second Transition	xxx	xxxx		xxx	xx	xxx	xxxx		xxx	xx	xx	xxx	xxxx	x	xxx	xx	xxxx	xx	xx	xxxx	xxxx	
Locus of Third Transition	xxxx	x	x	xxx	xxx	xxxx	x	x	xxx	xxx	xxxx	xxxx	x		xxx	xxx	xxxx	xxx	xx	xxxx	x	xxxx
Noise Frequency	x	xx	xxx	xx	xxxx							xx	xx	xx	xx	xxx						
Noise Intensity			x		x							x	x	x	x							
Transition Tempo	xx	x		x	x	xx	x		x	x	xx	x	x	x	x							
Closed Resonances						x	x	x	x	x								x	x	x	x	xx
			A					B						C					D			

FIGURE 2

Distinctive power of the acoustic cues of PLACE OF ARTICULATION in English.

A. Opposition of voiced stop consonants to each other.
B. Opposition of nasal consonants to each other.
C. Opposition of fricative consonants to each other.
D. Opposition of resonant consonants to each other.

("Ge" indicates "g" before spread vowels; "ga" indicates "g" before all other vowels.)

		p/m	t/m	ke/m	ka/m	b/m	d/m	ge/m	ga/m	f/m	θ/m	s/m	ʃ/m	v/m	ð/m	z/m	ʒ/m	n/m	ŋe/m	ŋa/m	w/m	j/m	r/m	l/m
Place of Articulation	Locus of Second Transition		$\overset{x}{x}{}^{x}$	$\overset{xx}{xx}$			$\overset{x}{x}{}^{x}$	$\overset{xx}{xx}$			$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{xx}{xx}$		$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{xx}{xx}$	$\overset{x}{x}{}^{x}$	$\overset{xx}{xx}$			$\overset{xx}{xx}$	x_x	x_x
	Locus of Third Transition		$\overset{xx}{xx}$	x	x		$\overset{xx}{xx}$	x	x		$\overset{xx}{xx}$	$\overset{xx}{xx}$	x		$\overset{xx}{xx}$	$\overset{xx}{xx}$	x	$\overset{xx}{xx}$	x	x		$\overset{xx}{xx}$	$\overset{xx}{xx}$	x_x
	Frequency of Noise		x_x	x_x	$\overset{xx}{xx}$		x	x	$\overset{x}{x}{}^{x}$		x	$\overset{xx}{xx}$	$\overset{xx}{xx}$			$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$							
	Intensity of Noise		x	x	x						x	x	x											
	Transition Tempo		x_x	x			x_x	x			x_x	x	x		x_x	x	x	x_x	x			x	x	x
	Closed Resonances		x	x	x		x	x	x		x	x	x	x	x	x	x	x	x	x				
Manner of Articulation	Continuity / Discontinuity									$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$				$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$
	Locus of First Transition	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$				$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$
	Transition Tempo									x_x	x_x	x_x	x_x	x_x	x_x	x_x	x_x				$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$
	Closed Resonances	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$				$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{xx}{xx}$	$\overset{xx}{xx}$
	Duration of Noise	x_x	x_x	x_x	$\overset{xx}{xx}$	x	x	x	$\overset{x}{x}{}^{x}$	x	x	$\overset{xx}{xx}$	$\overset{xx}{xx}$			$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$							
Voicing	Fundamental	x	x	x	x					x	x	x	x											
	Aspiration	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$					x	x	x	x											
	Intensity of First Transition	x_x	x_x	x_x	x_x					x_x	x_x	x_x	x_x											
	Intensity of Noise	x_x	x_x	x_x	x_x					x	x	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$											
	Closed Resonances	x_x	x_x	x_x	x_x					$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$	$\overset{x}{x}{}^{x}$											
	Duration of Hold	x_x	x_x	x_x	x_x					x_x	x_x	x_x	x_x											
	Duration of Following Vowel	x_x	x_x	x_x	x_x					x_x	x_x	x_x	x_x											

FIGURE 3

Distinctive power of the acoustic cues of PLACE OF ARTICULATION, MANNER OF ARTICULATION and VOICING in English. Opposition of a single consonant to all the others.

	ɛ̃/i	ɛ̃/e	ɛ̃/ɛ	ɛ̃/y	ɛ̃/ø	ɛ̃/œ	ɛ̃/u	ɛ̃/o	ɛ̃/ɔ	ɛ̃/a	ɛ̃/ɑ	ɛ̃/œ̃	ɛ̃/ɑ̃	ɛ̃/ɔ̃
Frequency of First Formant	xxx	xx		xxx	xx		xxx	xx		xx	xx			
Frequency of Second Formant	xxx	xx			xx	xx	xxx	xxx	xxx	xx	xx	xx	xx	xxx
Frequency of Third Formant	x													
Intensity of First Formant	xxx	xx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx			
Nasal Formant (250 cps)	x	x	x	x	x	x	x	x	x	x	x			
Vowel Duration	x	x	x	x	x	x	x	x	x	x	x			

FIGURE 4

Distinctive power of six acoustic cues of French VOWELS.
Opposition of a single vowel to all the others.

	p	t	ke	ka	b	d	ge	ga	f	θ	s	ʃ	v	ð	z	ʒ	m	n	ŋe	ŋa	w	j	r	l
p		x	x	xx xx	x	x	x	xx x	x	x	xx xx	xx xx	x	x	xx x	xx x	x	x	x	x	x	x	x	x
t	x		x	xx xx	x	x	x	xx x	x	x	xx xx	xx xx	x	x	xx x	xx x	x	x	x	x	x	x	x	x
ke	x	x			x	x	x	x	x	x	xx xx	xx xx	x	x	xx x	xx x	x	x	x	x	x	x	x	x
ka	xx xx	xx xx			xx x	xx x	x	x	xx x	xx x	xx xx	xx xx	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x
b	x	x	x	xx x				xx x	x	x	xx x	xx x			xx x	xx x								
d	x	x	x	xx x				xx x	x	x	xx x	xx x			xx x	xx x								
ge	x	x	x	x					x	x	xx x	xx x			xx x	xx x								
ga	xx x	xx x	x	x	xx x	xx x			xx x	xx x	xx x	xx x			xx x	xx x	x_x	x_x	x_x	x_x	x_x	x_x	x_x	x_x
f	x	x	x	xx x	x	x	x	xx x			xx x	xx x			xx x	xx x	x	x	x	x	x	x	x	x
θ	x	x	x	xx x	x	x	x	xx x			xx x	xx x			xx x	xx x	x	x	x	x	x	x	x	x
s	xx xx	xx xx	xx xx	xx xx	xx x	xx x	xx x	xx x	xx x	xx x		xx xx	xx x	xx x	x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x
ʃ	xx xx	xx xx	xx xx	xx xx	xx x	xx x	xx x	xx x	xx x	xx x	xx xx		xx x	xx x	xx x	x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x
v	x	x	x	xx x							xx x	xx x			x_x	x_x								
ð	x	x	x	xx x							xx x	xx x			x_x	x_x								
z	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	x	xx x	x_x	x_x		xx x	x_x	x_x	x_x	x_x	x_x	x_x	x_x	x_x
ʒ	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	xx x	x	x_x	x_x	xx x		x_x	x_x	x_x	x_x	x_x	x_x	x_x	x_x
m	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
n	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
ŋe	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
ŋa	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
w	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
j	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
r	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								
l	x	x	x	xx x				x_x	x	x	xx x	xx x			x_x	x_x								

FIGURE 5

Distinctive power of a single acoustic cue, that of NOISE (burst or friction noise) -- cue of Manner and of Place -- for all oppositions of two consonants in English.

	p	t	ke	ka	b	d	ge	ga	f	θ	s	ʃ	v	ð	z	ʒ	m	n	ŋe	ŋa	w	j	r	l
p		xx x	xx xx			xx x	xx xx		x	xx x	xx x	xx xx	x	xx x	xx x	xx xx		xx x	xx xx		x	xx xx	x_x	x_x
t	xx x		xx x	x_x	xx x		xx x	x_x	xx x	x_x	x_x	x_x	xx x	x_x	x_x	x_x	xx x		xx x	x_x	xx x	xx x	x_x	x_x
ke	xx xx	xx x			xx xx	xx x			xx xx	xx xx	xx xx	xx x	xx xx	xx xx	xx xx	xx x	xx xx	xx x			xx xx	x	xx xx	xx xx
ka		x_x				x_x				x_x	x_x	x_x		x_x	x_x	x_x		x_x			x	x_x	x_x	x_x
b		xx x	xx xx			xx x	xx xx		x	xx x	xx x	xx xx	x	xx x	xx x	xx xx		xx x	xx xx		x	xx xx	x_x	x_x
d	xx x		xx x	x_x	xx x		xx x	x_x	xx x	x_x	x_x	x_x	xx x	x_x	x_x	x_x	xx x		xx x	x_x	xx x	xx x	x_x	x_x
ge	xx xx	xx x			xx xx	xx x			xx xx	xx xx	xx xx	xx x	xx xx	xx xx	xx xx	xx x	xx xx	xx x			xx xx	x	xx xx	xx xx
ga		x_x				x_x				x_x	x_x	x_x		x_x	x_x	x_x		x_x			x	x_x	x_x	x_x
f	x	xx x	xx xx		x	xx x	xx xx			xx x	xx x	xx xx		xx x	xx x	xx xx	x	xx x	xx xx		x	xx xx	x_x	x_x
θ	xx x	x_x	xx xx	x_x	xx x	x_x	xx xx	x_x	xx x		x	x_x	xx x		x	x_x	xx x	x_x	xx xx	x_x	x_x	xx x	x	x
s	xx x	x_x	xx xx	x_x	xx x	x_x	xx xx	x_x	xx x	x		x	xx x	x		x	xx xx	x_x	xx xx	x_x	xx x	xx x	x_x	x_x
ʃ	xx xx	x_x	xx x	x_x	xx xx	x_x	xx x	x_x	xx xx	x_x	x		xx xx	x_x	x		xx xx	x_x	x_x	x_x	xx xx	x_x	x_x	x_x
v	x	xx x	xx xx		x	xx x	xx xx			xx x	xx x	xx xx		xx x	xx x	xx xx	x	xx x	xx xx		x	xx xx	x_x	x_x
ð	xx x	x_x	xx xx	x_x	xx x	x_x	xx xx	x_x	xx x		x	x_x	xx x		x	x_x	xx x	x_x	xx xx	x_x	x_x	xx x	x	x
z	xx x	x_x	xx xx	x_x	xx x	x_x	xx xx	x_x	xx x	x		x	xx x	x		x	xx xx	x_x	xx xx	x_x	xx x	xx x	x_x	x_x
ʒ	xx xx	x_x	xx x	x_x	xx xx	x_x	xx x	x_x	xx xx	x_x	x		xx xx	x_x	x		xx xx	x_x	x_x	x_x	xx xx	x_x	x_x	x_x
m		xx x	xx xx			xx x	xx xx		x	xx x	xx xx	xx xx	x	xx x	xx xx	xx xx		xx x	xx xx		x	xx xx	x_x	x_x
n	xx x		xx x	x_x	xx x		xx x	x_x	xx x	x_x	x_x	x_x	xx x	x_x	x_x	x_x	xx x		xx x	x_x	xx xx	xx x	x_x	x_x
ŋe	xx xx	xx x			xx xx	xx x			xx xx	xx xx	xx xx	x_x	xx xx	xx xx	xx xx	x_x	xx xx	xx x			xx xx	x_x	xx xx	xx xx
ŋa		x_x				x_x				x_x	x_x	x_x		x_x	x_x	x_x		x_x			x	x_x	x_x	x_x
w	x	xx x	xx xx	x	x	xx x	xx xx	x	x	x_x	xx x	xx xx	x	x_x	xx x	xx xx	x	xx xx	xx xx	x		xx xx	x_x	x_x
j	xx xx	xx x	x	x_x	xx xx	xx x	x	x_x	xx xx	xx x	xx x	x_x	xx xx	xx x	xx x	x_x	xx xx	xx x	x_x	x_x	xx xx		xx xx	xx xx
r	x_x	x_x	xx xx	x_x	x_x	x_x	xx xx	x_x	x_x	x	x_x	x_x	x_x	x	x_x	x_x	x_x	x_x	xx xx	x_x	x_x	xx xx		
l	x_x	x_x	xx xx	x_x	x_x	x_x	xx xx	x_x	x_x	x	x_x	x_x	x_x	x	x_x	x_x	x_x	x_x	xx xx	x_x	x_x	xx xx		

FIGURE 6

Distinctive power of a single acoustic cue, that of TRANSITION OF THE SECOND FORMANT -- cue of Place and Manner -- for the oppositions of two consonants in English.

	i	e	ɛ	y	ø	œ	u	o	ɔ	a	ɑ	ɛ̃	œ̃	ɑ̃	ɔ̃
i		xx	xxx		xx	xxx		xx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx
e	xx		xx	xx		xx	xx		xx	xxx	xxx	xx	xx	xx	xx
ɛ	xxx	xx		xxx	xx		xxx	xx		xx	xx				
y		xx	xxx		xx	xxx		xx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx
ø	xx		xx	xx		xx	xx		xx	xxx	xxx	xx	xx	xx	xx
œ	xxx	xx		xxx	xx		xxx	xx		xx	xx				
u		xx	xxx		xx	xxx				xxxx	xxxx	xxx	xxx	xxx	xxx
o	xx		xx	xx		xx				xxx	xxx	xx	xx	xx	xx
ɔ	xxx	xx		xxx	xx					xx	xx				
a	xxxx	xxx	xx	xxxx	xxx	xx	xxxx	xxx	xx			xx	xx	xx	xx
ɑ	xxxx	xxx	xx	xxxx	xxx	xx	xxxx	xxx	xx			xx	xx	xx	xx
ɛ̃	xxx	xx		xxx	xx		xxx	xx		xx	xx				
œ̃	xxx	xx		xxx	xx		xxx	xx		xx	xx				
ɑ̃	xxx	xx		xxx	xx		xxx	xx		xx	xx				
ɔ̃	xxx	xx		xxx	xx		xxx	xx		xx	xx				

FIGURE 7

Distinctive power of a single acoustic cue, the FREQUENCY OF THE FIRST FORMANT, for all vowel oppositions in French.

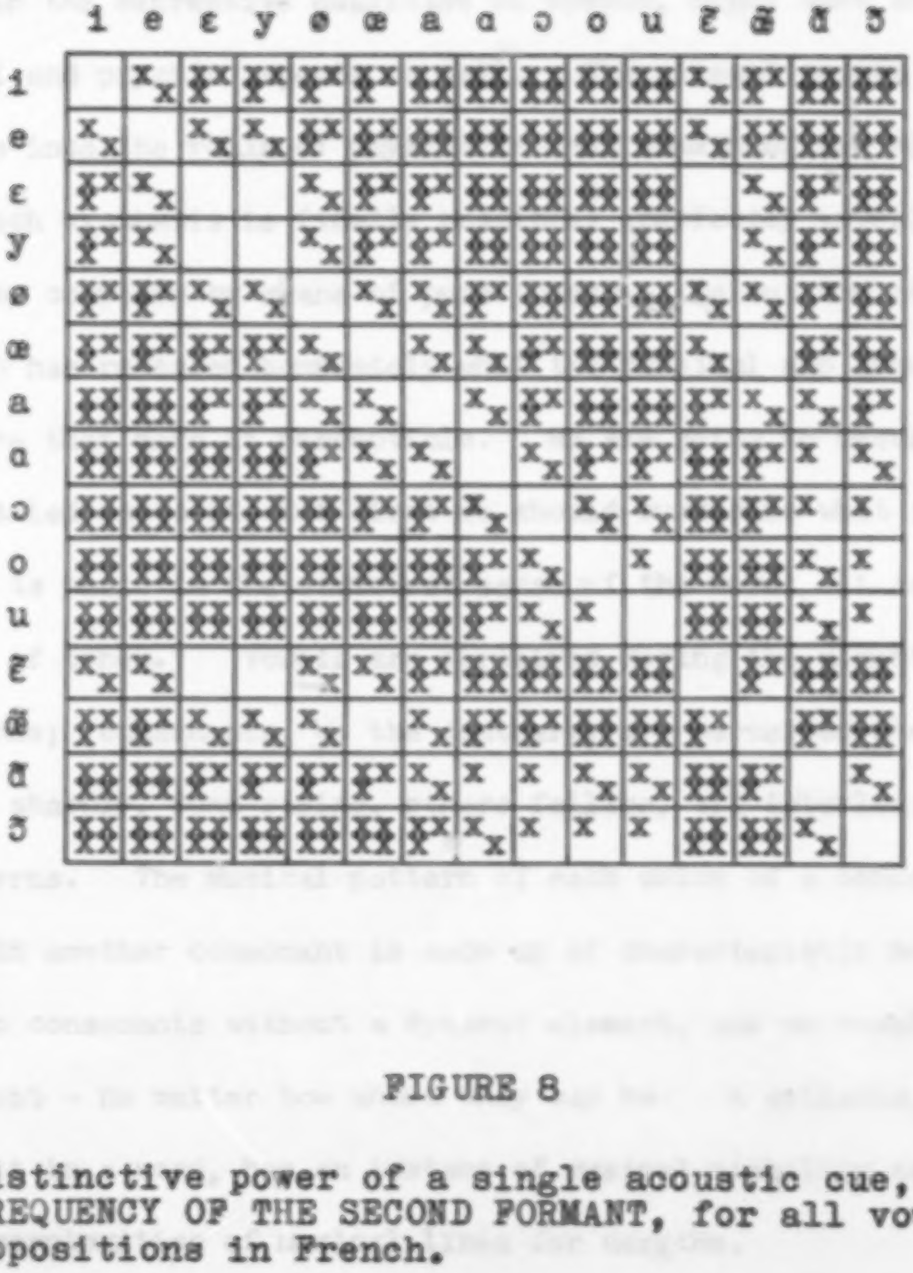

i e ɛ y ø œ a ɑ ɔ o u ɛ̃ œ̃ ɑ̃ ɔ̃
i e ɛ y ø œ a ɑ ɔ o u ɛ̃ œ̃ ɑ̃ ɔ̃

FIGURE 8

Distinctive power of a single acoustic cue, the FREQUENCY OF THE SECOND FORMANT, for all vowel oppositions in French.

APPENDIX III

THE ACOUSTIC SYNTHESIS OF SPEECH

(Translated from the French)

This scientific description is written for literary men who, being already familiar with the expressive qualities of speech, might want to inquire into its physical and psychic aspects as well. The alleged reason for proposing this venture into the realm of physics is that something new has been achieved there: speech synthesis is finally possible; perfectly intelligible artificial speech can be composed by means of pure tones -- the outcome of centuries of research which has revealed completely what the physical and acoustic elements of speech are that make it perceptible. We are going to describe this synthesis and what led us to it, but first we should summarize what it has taught us.

Speech is music in the strictest sense of the word: it is "a pleasing arrangement of tones." Vowels are perceived during the moments of stability of these tones; consonants, on the contrary, are perceived when these tones move rather sharply, some rising, others falling, all interlacing in infinitely varied patterns. The musical pattern of each union of a consonant with a vowel or with another consonant is made up of characteristic movements. Thus, there are no consonants without a dynamic element, and no vowels without a stable element - no matter how short they may be. A syllable, the smallest rhythmic unit in a word, has an instant of musical stability as its center and moving complexities of musical lines for margins.

Well may the poet rejoice. The medium through which he works to convey his feelings is truly lyrical. But the lyre which accompanies his words and that of the words themselves are of very different types. These two lyres play upon incompatible scales; they make use of musical intervals of opposed progressions. Music-music, that is music, as such, employs intervals, half-tones, which are (subjectively) all equal to the ear, but they differ in the number of vibrations -- objectively these intervals are in an increasing geometric progression, from low notes to high notes. Just the opposite is true of speech-music, which uses intervals that are all equal objectively (intervals of the same number of vibrations between the low notes as between the high notes), but that are unequal subjectively. The series of tones which speech uses is known as a harmonic series. Any schoolboy knows that the notes of such a series are multiples of the fundamental (first harmonic) by whole numbers: if the fundamental is 150 cycles per second, the 2nd harmonic has a frequency of 300, the 3rd of 450, the 4th of 600, etc.. But, for the ear this means that the 2nd harmonic is one octave above the fundamental (first harmonic), the 3rd harmonic, one fifth above the second, the 4th harmonic one fourth above the 3rd, the 5th one major third above the 4th, the 6th one minor third above the 5th, and so on. The intervals of a harmonic series become smaller and smaller as the scale climbs, so that the interval which separates the 49th harmonic from the 50th is only about one third of a tone. It is by means of this strange range of unequal intervals that our mouth "composes" the musical chords of vowels and consonants. These are neither tones nor half-tones, but an infinite variety of intervals, all slightly different from the next one. At 12,000 cycles, the upper limit of frequencies used in speech, the intervals attain an eighth of a tone for a man's voice and a quarter tone for a woman's voice (whose fundamental is on an average an octave higher than that of a man). Let us note in passing that women use half as many harmonics as do men; but not to be outdone as to expenditure, women produce twice as many vibrations per second!

Here then is a first distinction between speech-music and that of man-made musical instruments. There are others. In order to understand them, let us start by observing how the mouth produces the sounds of a language.

The mouth has all the attributes of a musical instrument. These attributes are essentially a sound source and a system of resonance that gives the sound coming from this source a particular character. If a saxophone and a clarinet, each hitting the same A-4, produce a different timbre -- the saxophone timbre, the clarinet timbre -- it is because their resonance chambers are different. In both cases the source, put into action by the musician's lips produces a fundamental, the A-4, and a rich series of harmonics of this fundamental. However, before reaching the ear, this series of harmonics is resonated, or more precisely, filtered, by the resonance cavities of the instrument, which "pass" those harmonic components whose notes are close enough to the notes of the natural resonance of the instrument's cavities. The frequency ranges of resonance in each instrument are called its "formants." The formants of a saxophone are different from those of a clarinet -- the saxophone chooses different components among those of the source -- because the volume and form of its resonance chambers are different.

The difference in color (timbre) between two vowels is analogous to that which distinguishes two musical instruments. The color of every vowel is characterized by its formants. In order to produce vowels, the mouth uses a source, the vibrating vocal cords, and a system of resonance, the buccal cavities (to be more exact, the buccal, pharyngeal and nasal cavities). For an i as for an a, the vocal cords emit a rich series of harmonics, exactly the same in both cases if the two vowels have the same fundamental note. But, in producing i the buccal cavities take on a shape and volume appropriate to this vowel, which permits them to choose (to filter, to resonate) the harmonic components of the i and to eliminate all others. In producing an a, the mouth assumes another shape; its cavities assume other volumes, and the components which are resonated are different. Let us take a specific example. For an i the distinctive components are grouped in two frequency regions, one very low on the vowel scale and one very high; for an a they are grouped in two middle regions rather close to each other. These frequency regions are the formants. Each formant can include several harmonic components because the resonance cavities in the mouth are not narrowly tuned; they resonate harmonics within a margin of 200 or 300 cycles. It is possible to hear the approximate note of the upper formant of an i or an a by whispering these vowels, and to hear their lower note by tapping the upper neck with a pencil, thereby generating a resonance in the region of the pharynx. In whispered voice, no harmonics (periodic sound) are used. Formants are composed simply of noise (non-periodic sound). The source is still in the vocal cords, but instead of vibrating, the cords remain immobile, forming a narrow slit through which the breath escapes under pressure and produces a friction noise of a wide frequency-range. While going through the buccal cavities, this noise is filtered, resonated, exactly as harmonics would be, and the mouth radiates several narrow bands of noise, the whispered formants, which have the same frequencies as if the formants were composed of harmonics.

To complete this description, we must mention the relationship between the frequency of formants, which characterizes vowel color, and the fundamental frequency of the voice, which characterizes the melody in singing or the intonation in speech. These two phenomena are independent of one another. The frequency of the formants depends upon the volume and shape of the cavities, which amplify all noises or harmonics whose frequency happens to coincide with theirs; whereas, voice pitch only depends upon the speed at which the vocal cords vibrate. To verify this independence of source pitch and resonated formant pitch, it suffices to observe that one can change vowels without modifying the note of the voice (i-a-o can be spoken or sung on the same D-2), just as one can change notes without modifying the color of the vowel that is being spoken or sung.

Now let us return to what distinguishes music-music from speech-music. We have seen that both types of sound use musical scales; the musical scale of music, however, uses intervals which sound equal to the ear but which are unequal in cycles per second, whereas the musical scale of speech uses intervals unequal to the ear but which are equal in cycles per second.

Another difference is seen in the complexity of the tones of these scales. Every note of a musical instrument possesses a fundamental and certain harmonics of this fundamental; therefore, notes of an instrumental scale are complex tones. Notes of the scale of harmonics which compose a vowel are pure tones, that is, simple sine waves. Thus, the chord that someone strikes on the piano combines notes each of which forms a group of complex tones; it is a chord of fundamentals in which each fundamental is accompanied by its harmonic components. The chord which makes a vowel audible only combines pure tones; it is a chord of components chosen from among the harmonics of a fundamental. Moreover, the fundamental, which is actually the lowest harmonic in the series, is not even necessary to make a vowel audible; it is sufficient that the harmonics appropriate to this vowel be present. Thus we can see why the notes of a piano cannot compose a vowel -- piano notes are complex notes with many harmonics. On the other hand, excellent vowels can be composed by striking tuning forks, or by whistling if several persons whistle the right notes in unison. But sets of pure tones can be more efficiently produced by various electronic means.

We are now getting close to actual speech synthesis. Since mathematical analysis had shown that all speech sounds could be broken down into a certain number of pure tones harmonically linked to one another, it seemed that if one recombined pure tones produced electronically in a harmonic relation one should obtain the sounds of vowels. This was done and the result was positive. Through experimentation with pure tones, it was then discovered that the harmonics of the first three formants, the three lowest ones were enough to distinguish one vowel from another, and that the upper formants serve only in distinguishing one voice from another. The frequency of upper formants is known by electronic analysis; there can be as many as six or eight of them. The three low formants have a linguistic function, and the high formants have a social or aesthetic function.

To avoid confusion, we have, so far, only discussed vowels. Their synthesis is quite simple. With three pure tones at the frequency of the first three formants, the color of a vowel can already be recognized. With eight or ten pure tones a perfectly human vowel can be obtained. But the synthesis of consonants presents infinitely more complex problems.

Consonants are perceived by noise, or harmonic murmurs, during maximum buccal constriction, and by rapid frequency changes in the formants while the mouth is opening or closing. The noises are the bursts of stop consonants as in pa, ta, ba, or the frictions of fricative consonants as in fa, cha, za. These bursts and frictions are composed of very complex, non-periodic waves which are distinguished from each other by their frequency, their intensity, their band width and their duration (the friction of an s is higher than that of a ch, that of an f is weaker and of wider frequency than that of ch, the explosion of a t is shorter than the friction of an s even though they both have the same pitch, etc.). The murmurs which are produced during the most close part of consonants are those which start syllables such as ma, na, la, rat, hier, oui, huit. Their formants are analogous to those of vowels, but with less force. These are "murmurs" in comparison with the clear sounds of vowels which follow them.

Until very recently it was thought that consonants were perceived by these noises and murmurs. All the phonetic manuals defined consonants as noises accompanied or unaccompanied by vibrations of the vocal cords. This was incorrect. The research of the past ten years, by means of the speech synthesizers we are about to describe, has demonstrated that the perception of consonants depends essentially upon rapid frequency changes of the formants -- changes which reflect the rapid movements of opening or closing which are articulated in moving from the closed position of the consonant to the open position of the vowel, as in ba, or doing the opposite as in ab. These transitory portions between closing and opening were always neglected and were attributed to vowels. Such an interpretation was unjustified. These changing portions of syllables, these "transitions," do not contribute anything to the perception of vowels, but they are indispensable to the perception of consonants. The role of noises and murmurs is not entirely negligible but what permits us to distinguish ba from da, ma from na, la from ra, is mainly the transitions of formants. The synthesizers have demonstrated this clearly. One might have predicted as much because noises and murmurs do not carry, that is, they are only heard if the speaker is very close to the hearer -- at a distance of several yards these noises and murmurs fade out. On the other hand, the transitions of formants have the intensity of vowels -- from the point of view of perception at a distance, they are moving vowels and they carry as far as vowels. It is probable that, not being able to hear consonant noises and murmurs at a distance because of their weakness, men have become accustomed to hearing them by their transitions, that is to say, by substituting vowel movement for consonant noise. At least this is the psychologists' explanation.

To succeed in synthesizing consonants, it was, therefore, necessary to discover the laws of transitions of formants for the perception of syllables. This is the research to which we dedicated the years from 1950 to 1960. It was done by means of synthesizers. For speech synthesizers are, above all, research instruments. They make it possible to isolate the acoustic elements that are combined in syllables, to vary the acoustic dimensions of each element while keeping the other elements fixed, to judge by ear the effects of these modifications, and thus to discover the relative importance of each element in overall comprehension. These manipulations of synthetic speech make it possible to discover the optimum conditions for the perception of each element -- conditions which should lead to the creation of more intelligible speech, more resistant to noise, and more pleasing to the ear than natural speech (for which all mouths are not equally endowed). Moreover, the speech synthesizer is not limited to linguistic factors; it can also be used in studying aesthetic effects, since it controls voice quality by the upper formants, it is capable of producing a spoken or sung voice (with synthetic vibrato) of a color superior to that of the human voice.

The reason attempts to synthesize real sentences had failed until recently is precisely that the transitions had been neglected. Noises or murmurs of consonants had been juxtaposed to periodic vowel sounds without being joined by formant transitions, and the results had always been unintelligible. It was only when the essential role of rapid frequency changes in transitory movements was understood that we were on the right track. Speech is not made up of static portions but of a dynamic continuity. In order to synthesize speech one needs a dynamic machine capable of reproducing in time all the frequency and intensity changes which the transitory formants and even noise might undergo.

We will now describe the tone-wheel synthesizer which was constructed in our laboratories in Boulder and then transported to the RESEARCH LABORATORY OF EXPERIMENTAL PHONETICS which is under our direction at Santa Barbara.

The 50 pure tones of this synthesizer are produced by means of a tone wheel, which is the heart of the machine. This wheel, produced photographically by variable light intensities, contains 50 circles, on which variable densities of opacity alternate. On the smallest circle, the opacity alternates four times from dark to light. When the wheel turns 30 times per second, a ray of light which passes through the wheel is modulated 120 times per second by the 4 dark segments of this first circle. This circle is responsible for the production of the fundamental tone. The second circle, having 8 dark segments, modulates the light 240 times per second. The third, having 12 segments, modulates it 360 times, and so on up to the 50th, which, having 200 dark segments, modulates the light 6000 times per second. The modulations of light by the dark segments are not abrupt; they are gradual so that they may correspond to the variations in amplitude of a sinusoidal wave, without which complex tones and non-pure tones would be obtained. Therefore, the passage from dark to light on each circle must be by gradual variation.

Now let us look at figure I. The light of a powerful arc (A), condensed by a lens (B), passes through the tone wheel (C), and comes out again at the right transformed into 50 beams of modulated light of different frequencies ranging from 120 cycles to 6000 cycles per second. These 50 beams are picked up by a lens (D) and are condensed on a mirror at angle (E), which projects them vertically onto the transparent plastic band (H), where the notes of the speech-music scale are written. These notes are painted in opaque white so as to reflect the beams of light. Where there is no paint, the light passes through the band and finally dies out in the dark chamber (F) which is below; wherever there is any paint, the light is reflected upwards. The beams of reflected light are collected by transparent lucite fingers at the bottom of the box (G) and are guided toward a photo-electric cell which transforms the modulations of the beams of light into sound waves. Finally these sound waves pass through an amplifier (J), which sets off the vibrations of a loud speaker (K). The transparent band which carries the musical patterns of speech moves from right to left, like the roll of a player piano, and makes the speech patterns pass under the 50 beams of light which, sweeping these musical patterns, select at each instant the pure tones which should be played. Thus an easily manipulated element, the beams of light, is used to choose the desirable tones and, lastly, the light vibrations are transformed into sound vibrations, so they can be heard. The chords of continually changing pure tones which come from the speaker are understood as human speech.

We mentioned at the beginning of this discussion that in order to characterize all complex speech sound, two forces are involved -- a source and a resonance. To produce speech, man was endowed with a source of pure tones by means of the vocal cords, and a selective resonance of those pure tones by means of resonance cavities in the mouth. In the same way our machine operates by a source of pure tones, the musical pattern, which is kept moving by means of the plastic band. The tone wheel plays the part of the vocal cords and the musical pattern that of the buccal resonators.

Figure 2 shows the musical notes of two lines of Apollinaire:

L'anémone et l'ancolie
ont poussé dans le jardin

as they are painted on the plastic band to be passed under the beams of light. The musical range goes from 120 cycles for the lowest notes of the vowel i (at bottom right) to 6000 for the highest notes of the noise of the s.

It will be noted that tones are grouped in three bands which curve from left to right. These are the transitory movements of the first three formants. These formants reflect the articulatory movements of the tongue and other buccal organs, whose function is to change constantly the form and volume of the buccal cavities. As to the physics of sound, the first line of poetry is more periodic than the second; apart from the small burst of noise at the beginning of the s, it is entirely made up of harmonics. The second line is more turbulent: the burst noise of the p and the friction noises of the s and the j are made up of non-periodic sound. The subjective impression does not seem to contradict the objective data as given: everyone finds the first line more "musical" than the second. It is thus interesting to notice that the acoustic breaking down of speech into simple sine waves, or pure tones, renders the musical qualities visible. The poetic factors which relate to the sound can literally be seen on the speech patterns. And not only are the periodicity and the turbulence visible, but all the musical factors: the equal length of the syllables which characterizes French rhythm, the lengthening of the last syllable which realizes stress, the frequency of the second formant which is the most audible and best characterizes the vowels to the ear. We also note that the highest tones of the second formant are found in the i and the lowest ones in the ou. It can also be seen that what gives the muffled character to nasal vowels is a lack of balance of the intensities: if there is nasalization, the first formant is much weaker than the second (In synthesis, in order to nasalize a vowel it suffices to reduce sharply the intensity of the first formant without changing that of the other.) Finally, it can be seen why certain consonants are softer to the ear than others: the l, r, m, n sounds are composed entirely of harmonics, even during the murmur which is produced when the organs of articulation are the most close. For consonants such as p, k, s, ch, the portion produced while the organs are close is composed of unharmonic or turbulent sounds.

Examples could be multiplied, but let us stop there and say a few words in closing about the ways in which speech synthesis can be of use.

In the first place, it has permitted the discovery and cataloging of all the acoustic cues for the perception of English and satisfied scientific curiostiy by adding a brand new chapter to our knowledge in linguistics, as well as to our knowledge in physics. Some will say that this is abstract research without practical application. Far from that. Applications will be made in various fields. Let us cite a few of them.

If someone manages to perfect the typewriter which will automatically type what the author dictates into a microphone, it will be thanks to the knowledge of acoustic cues that we have discovered in the first ten years of this type of research.

The doctor who learns to read acoustic speech patterns will be able to diagnose certain nervous or mental diseases. He will also be able to study the effect of surgery in the throat, the mouth, or the nose, by comparing the acoustic speech patterns before and after the operation.

The professor of language or speech who knows how to read these acoustic patterns can find on them all of his student's articulation errors, even certain mistakes he could not hear before seeing them.

After an airplane accident, the recording of the pilot's last words, often incomprehensible, can be deciphered by acoustic patterns.

By manipulation of synthetic speech, psychological experiments concerning the comprehension and learning of a language are being conducted. We have already learned, for example, that language sounds are not understood directly by the brain, as they come from the ear; but, instead they are understood indirectly after reference to a learned articulatory habit.

Lastly, with the aid of the Department of Education in Washington, we are presently applying on a large scale the research technique of synthetic speech to the objective analysis of the articulatory differences between the most widely taught languages: English, French, Spanish, German. By the technique of synthesis, we are able to isolate 20 to 30 differences between two languages, to judge by ear the degree of importance of each of these differences, to give, for example, to an English sentence around 20 different degrees of French accent by simply pushing the right buttons on the synthesizer or by making appropriate changes on the musical patterns. A thorough knowledge of the distinctions between two languages will make an important contribution to the improvement of language teaching.

FIGURE CAPTIONS

Figure 1. Schematic representation of a speech synthesizer.

Figure 2. The notes of musical patterns for the synthesis of two lines of Apollinaire: L'anémone et l'ancolie -- Ont poussé dans le jardin. The range of frequencies is from 120 to 6000 per second, from bottom to top.

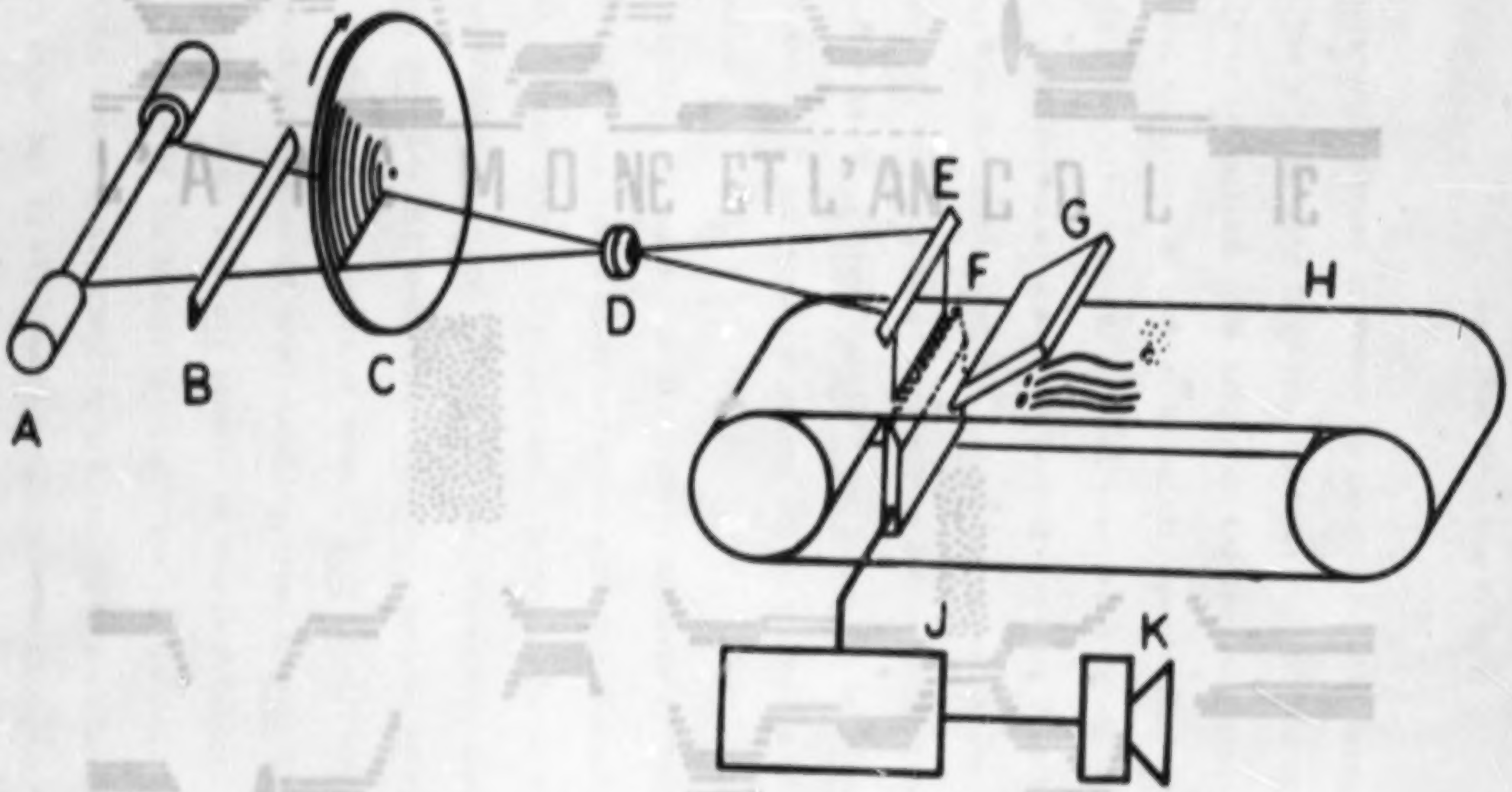

A
B
C
D
E
F
G
H
J
K

Figure 1

Diagram of Pattern Playback

APPENDIX IV

VOWEL NASALITY IN FRENCH AND ENGLISH [1]

(translated from the French)

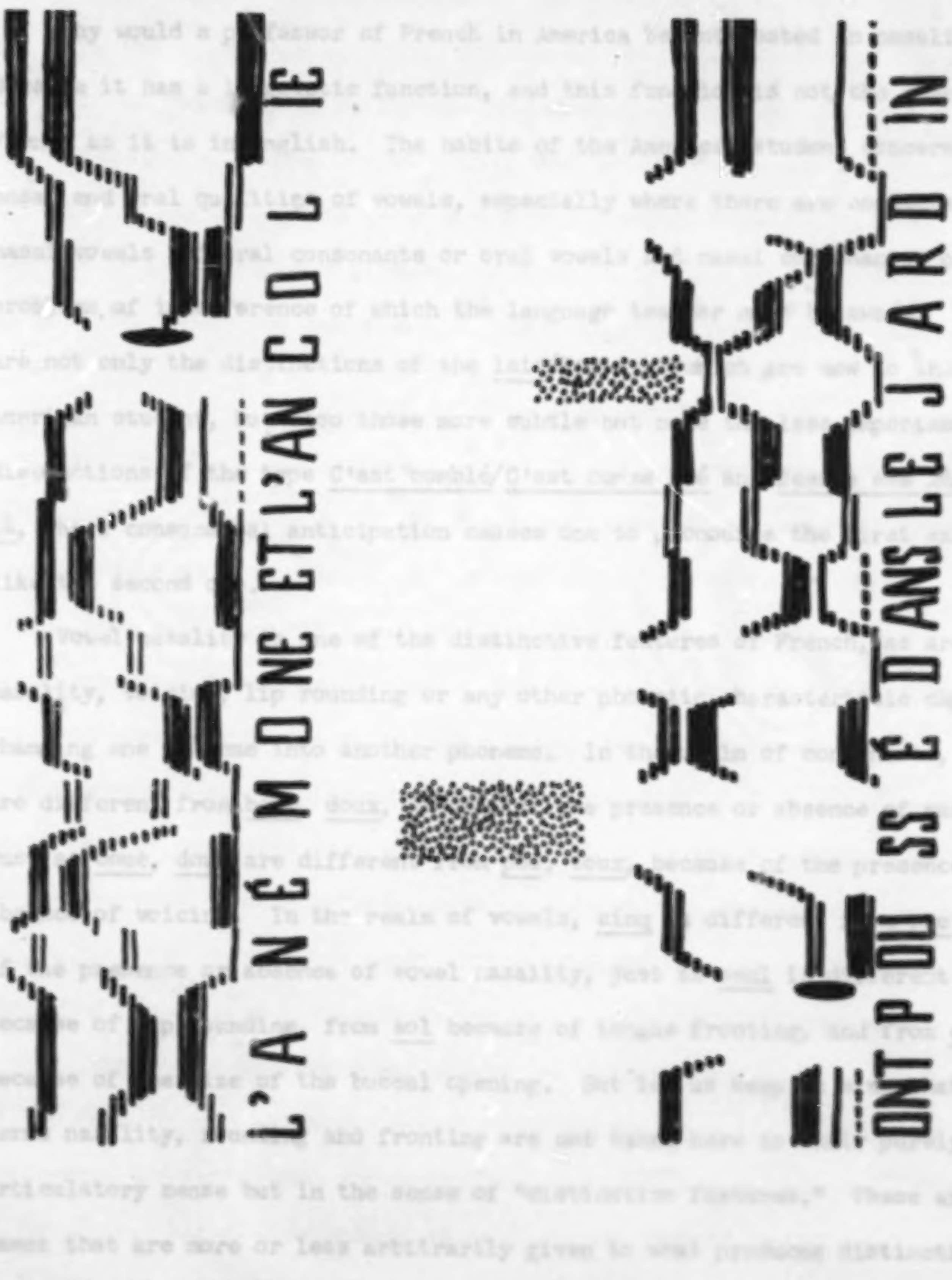

L'A N É M O NE ET L'AN C O L IE
ONT P OU SS É D ANS LE J A R D IN

Figure 2
Hand-painted speech patterns for synthesizing two lines of poetry by Apollinaire:
L'anémone et l'ancolie, Ont poussé dans le jardin.
The frequency range is from 120 cycles per second (bottom) to 6000 cycles per second (top).

Why would a professor of French in America be interested in nasality? Because it has a [illegible] function, and this function is not the same in French as it is in English. The habits of the American student concerning nasal and oral qualities of vowels, especially where there are [illegible] of nasal vowels [illegible] oral consonants or oral vowels and nasal consonants, present problems of interference of which the language teacher must be aware. These are not only the distinctions of the [illegible] which are new to the American student, but also those more subtle but none the less important distinctions of the type *C'est tombé*/*C'est [illegible]* [illegible] *J'en ai là*, where consonantal anticipation causes one to pronounce the first example like the second one.

Vowel nasality is one of the distinctive features of French, as are [illegible] nasality, voicing, lip rounding or any other phonetic characteristic capable of changing one phoneme into another phoneme. In the realm of consonants, [illegible] are different from [illegible] *doux*, because of the presence or absence of nasality, just as [illegible], *doux* are different from [illegible], *toux*, because of the presence or absence of voicing. In the realm of vowels, *cinq* is different from [illegible] because of the presence or absence of vowel nasality, just as *seul* is different from [illegible] because of lip rounding, from *sol* because of tongue fronting, and from *cil* because of the size of the buccal opening. But let us keep in mind that the terms nasality, rounding and fronting are not taken here in their purely articulatory sense but in the sense of "distinctive features." These are the names that are more or less arbitrarily given to what produces distinctions like *sec*/*cinq*, *seul*/*sel*, *seul*/*sol*. Thus, the distinctive feature "nasality" actually

APPENDIX IV

VOWEL NASALITY IN FRENCH AND ENGLISH [1]

(translated from the French)

Why would a professor of French in America be interested in nasality? Because it has a linguistic function, and this function is not the same in French as it is in English. The habits of the American student concerning nasal and oral qualities of vowels, especially where there are contacts of nasal vowels and oral consonants or oral vowels and nasal consonants, present problems of interference of which the language teacher must be aware. These are not only the distinctions of the *laid*/*lin* type which are new to the American student, but also those more subtle but none the less important distinctions of the type *C'est comblé*/*C'est comme blé* and *Jeanne est là*/*J'en ai là*, where consonantal anticipation causes one to pronounce the first example like the second one.

Vowel nasality is one of the distinctive features of French, as are consonant nasality, voicing, lip rounding or any other phonetic characteristic capable of changing one phoneme into another phoneme. In the realm of consonants, *mou*, *nous* are different from *boue*, *doux*, because of the presence or absence of nasality, just as *boue*, *doux* are different from *pou*, *toux*, because of the presence or absence of voicing. In the realm of vowels, *cinq* is different from *sec* because of the presence or absence of vowel nasality, just as *seul* is different from *sel* because of lip rounding, from *sol* because of tongue fronting, and from *cil* because of the size of the buccal opening. But let us keep in mind that the terms nasality, rounding and fronting are not taken here in their purely articulatory sense but in the sense of "distinctive features." These are the names that are more or less arbitrarily given to what produces distinctions like *sec*/*cinq*, *seul*/*sel*, *seul*/*sol*. Thus, the distinctive feature "nasality" entails several phonetic differences other than physiological nasality -- we shall see later that /ɑ̃/ is longer, more closed, farther back than /ɑ/. Taken as a distinctive feature the term nasality then will include all the phonetic factors which contribute to the distinction of the *sec*/*cinq* type.

The subjective and objective aspects of vowel nasality which can be of interest in the teaching of French are numerous and varied. We propose to summarize them all here. It is a good time to do this because recent research by means of motion picture x-ray and by the manipulation of artificial speech has enabled us to take a step forward, to find the answers to some questions, and to form new hypotheses on others, in physiology as well as in acoustics.

THE PHYSIOLOGICAL ASPECT

In Figure 1, the articulation of nasal and oral vowels is taken from actual photographs. We see here for the first time tracings of x-rays of nasal vowels spoken naturally, articulated at a normal speed in sentences, and not held or posed, thus free from the artificiality and exaggeration that stopping the articulatory movement would entail. The four nasal vowels of this illustration are taken from the phrase *un bon pain blanc*, the articulation of which was photographed by motion picture x-ray at the speed of 24 frames per second. Since each frame is separated from the next by about 4 cs (centiseconds), we have chosen for each vowel the frame which showed the most pronounced lowering of the velum. [2] After filming the nasal articulation of five native French speakers, and observing their velar movements on the movie screen at different speeds which varied from one to 24 frames per second, we chose for Figure 1 the speaker who, of the five, seemed the most typical. Variations among the five speakers were so small that the x-ray tracings of Figure 1 may be considered as composite tracings of the five. All of them spoke, as natives, the type of French used in the North of France (as opposed to that of the Midi or South of France), but only three of them had learned the four nasal vowels as children; the other two speakers were naturally accustomed to only three of the nasals, those of lin, lent and long, and had learned the fourth, that of l'un, with the lips rounded, only as adults for teaching purposes. The subject of Figure 1 was not one of these last two.

Let us examine in detail the articulatory positions of Figure 1.

The first column shows the four nasal vowels of French at their point of maximum velar opening.

The second column shows the four oral vowels which correspond to the nasals when they are de-nasalized: since plein de-nasalizes into pleine, jeun into jeune, plan into plane and bon into bonne, the radiograms of column 2 are those of /ɛ/, /œ/,/a/ and /ɔ/, which have also been filmed within words, and are shown next to their corresponding nasals of /ɛ̃/, /œ̃/, /ɑ̃/ and /ɔ̃/. [3]

The third column of Figure 1 shows, for comparison, vowels from which the nasals originate: /ɛ̃/ sometimes comes from [ĩ], /œ̃/ from [ỹ], /ɔ̃/ from [õ] or [ũ], as we shall see in the section on the historical aspect.

The four vowels have two articulatory traits in common.

a. They all have the soft palate lowered in the form of an arc. This lowering opens the velic passage and apparently puts the velic cavity (small nasal cavity just above the soft palate) in direct communication with the pharyngeal cavity (the cavity whose upper limits are formed by the dorsal constriction of the tongue against the middle of the pharyngeal wall, and whose lower limits are the vocal cords and the opening of the esophagus). [4]

The arched form of the lowering of the soft palate that we observe in these four radiograms is characteristic of nasal VOWELS, but not of nasal consonants. Let us take a look at Figure 3, where we can observe the articulation of nasal consonants. In Sam and sank, said by an American, as in blâme, said by a Frenchman, the velar opening of the consonants is made not be a lowering of the velum, but by a lateral pleated folding, which creates an opening narrower than that of nasal vowels.

b. The four nasal vowels have about the same volume of pharyngeal cavity. 5 Since these equal volumes were attained as the result of a long evolution, we must suppose that vowel nasality is facilitated by an optimum volume of pharyngeal cavity which, once attained, stays fixed. Originally, in fact, the vowels which underwent nasalization by the spreading of nasality from a contiguous consonant like /m/ and /n/, could have a smaller pharyngeal cavity than today's nasals, like /a/ in the second column, or larger, like /i/, /y/, /u/, /o/ in the third column of radiograms.

Since these two traits are common to the four vowels which are judged subjectively as nasals, and are not common to those judged to be oral, we are justified in linking them with vowel nasality which distinguishes nasal vowels from oral vowels. The other articulatory attributes that can be observed in the first column are, on the contrary, those which make it possible to distinguish the four vowels from each other.

c. With respect to the tongue, /ɛ̃/ and /œ̃/ are clearly more fronted than /ɑ̃/ and /ɔ̃/ -- the backing of these last two is such that the tongue almost comes in contact with the lowered velum -- but /œ̃/ is distinguished from /ɛ̃/, as is /ɔ̃/ from /ɑ̃/, by the rounding of the lips. Although the position of the tongue plainly places the four vowels in two groups, one can establish a progression of gradual backing in four steps: /ɛ̃/, /œ̃/, /ɑ̃/, /ɔ̃/.

d. As for the lips, /œ̃/ and /ɔ̃/ are clearly more rounded than /ɛ̃/ and /ɑ̃/, but /ɔ̃/ is distinguished from /œ̃/, as /ɑ̃/ is from /ɛ̃/, by the drawing back of the tongue toward the pharyngeal wall. Although the shape of the lips plainly places the four vowels in two groups, one can establish a progression of gradual rounding in four steps: /ɛ̃/, /ɑ̃/, /œ̃/, /ɔ̃/.

e. As to the opening of the jaws, which can be observed from the separation of the teeth, /ɛ̃/ and /œ̃/ are more open than /ɑ̃/ and /ɔ̃/. A gradation in four steps is also possible here; it parallels the gradual backing and consequently follows the progression: /ɛ̃/, /œ̃/, /ɑ̃/, /ɔ̃/.

f. With respect to the cavities, since the four vowels have the same size pharyngeal cavity (back), it is obviously by the buccal cavity (front) that they are distinguished from each other. The note of the second formant depends mainly on these buccal cavities. Now, it is known that the note of resonance of a cavity depends essentially on two very simple factors: the larger the cavity and the smaller the openings, the lower the note. Let us apply this principle to the buccal cavities of the four nasal vowels such as we observe them in Figure 1.

The buccal cavity of /ɛ̃/ resonates the highest note of the four because, although the cavity is rather large (between the pharyngeal constriction and the lips),it is very open at its two extremities (the tongue is fronted and the lips are spread).

The buccal cavity of /œ̃/ resonates a little lower note because, while it is as large as that of /ɛ̃/ in volume and in back openings, it is clearly more closed in front due to the rounding of the lips.

The buccal cavity of /ɑ̃/ resonates a still lower note because, although the cavity is somewhat smaller than that of /œ̃/, it is much more closed at the back.

Finally the buccal cavity of /ɔ̃/ resonates an even lower note, the lowest of the four because, while it is large, it has a narrow opening at both ends -- not only at the back, like /ɑ̃/, but also at the front, owing to a very pronounced rounding of the lips.

The notes of resonance of the buccal cavity which are predicted above by the shape and size of the cavities are entirely confirmed by the spectographic analysis of Figure 2 that we discuss below (cf. acoustic aspect). They can also be verified very simply by ear, by whispering the four vowels in a quiet room, free from outside noises.

g. Finally let us compare in Figure 1 the articulatory position of oral vowels and that of nasal vowels.

The vowel /ɛ̃/ comes very close to /ɛ/ in its opening and spreading. It differs noticeably from the oral only by a more pronounced pharyngeal constriction which better isolates the back cavity. For the nasal vowel the vocal tract takes on slightly less the shape of a tube and more that of two separate spheric cavities.

In the same way /œ̃/ is fairly similar to /œ/ except that it has a more pronounced rounding and a more spherical pharyngeal cavity. Thus the two fronted nasal vowels correspond rather well to their oral counterparts. The same thing cannot be said of the two back nasals.

The vowel /ɑ̃/ clearly differs from /a/ in several ways. For /ɑ̃/ the jaws are closer, the lips slightly less rounded, the top of the tongue farther back, and since the constriction is higher up in the pharynx, the pharyngeal cavity is larger. That should result in a lowering of the two formants. The profile of /ɑ̃/, moreover, does not resemble that of any other vowel. The most one might say is that the buccal cavity of /ɑ̃/ would come close to that of /o/ if there were rounding; /ɑ̃/ approximates the tongue position of /o/ and the lip position of /ɑ/. Its pharyngeal cavity is larger but less closed than that of /ɔ/, and smaller and less closed than that of /o/.

The buccal cavity of /ɔ̃/ is very different from that of /ɔ/. For /ɔ̃/ the jaws are more closed and the lips more rounded than for /ɔ/. The mass of the tongue is much more retracted. The pharyngeal cavity is larger, but since at the same time it is more open, it could resonate at the same note as the pharyngeal cavity of /ɔ/. On the whole these two vowels vary much more in their buccal cavity than in their pharyngeal cavity.

If we compare /ɔ̃/ and /o/, we observe just the opposite: the buccal cavities are alike in the retracting of the tongue, the largeness of the cavity, the closeness of the teeth and the rounding of the lips, but the pharyngeal cavities differ considerably -- that of the /o/ is larger and more closed than that of the /ɔ̃/.

Thus it can be said that /ɔ̃/ comes close to having the buccal cavity of /o/ and the pharyngeal cavity of /ɔ/. We shall see farther on that this is confirmed by the spectograms, since /ɔ̃/ combines the second formant of /o/ with the first formant of /ɔ/, at least from the point of view of frequencies.

THE DISTINCTIVE ASPECT

a. On the traditional basis of labial, lingual and maxillary (opening of the jaws) positions, the DISTINCTIVE FEATURES which separate the four nasal vowels from each other are above all tongue fronting/backing and lip rounding/spreading. It seems justified to consider the height of the tongue or the opening of the jaws as redundant. It is in fact possible to change /ɛ̃/ to /œ̃/, or /ɑ̃/ to /ɔ̃/ solely by rounding the lips, and to change /ɛ̃/ to /ɑ̃/ or /œ̃/ to /ɔ̃/ solely by backing the tongue. On the other hand, only lowering the tongue or opening the jaws does not produce a well-defined change. Furthermore, we have noticed above that the lessening of the maxillary opening parallels the backing of the tongue in four steps from /ɛ̃/ to /ɔ̃/, which allows us to consider the maxillary closing as a secondary effect of the backing of the tongue.

Therefore, we can compare French nasal vowels to each other by only two distinctive features, and consider the others as more or less redundant.

	TONGUE	LIPS	NOTE OF RESONANCE
ɛ̃	front	spread	high
œ̃	front	rounded	mid-high
ɑ̃	back	spread	mid-low
ɔ̃	back	rounded	low

This description by labial and lingual distinctive features has the advantage of showing that [ɑ̃] is not distinguished from the other nasals by a larger degree of opening, as was believed on the basis of its presumed connection with the vowel /a/, a connection which is mostly based on spelling. But, the description is not entirely satisfactory because if /ɑ̃/ is backed, it is less sharply backed than /ɔ̃/, and if the lips are spread, they are less decidedly spread than for /ɛ̃/. Moreover this labial and lingual description has the curious effect of introducing into French a back-spread vowel, the only one of its kind, for the back-spread category is characteristic not of French but of English. So let us try other kinds of descriptions.

b. The distinctive features can also be established in a completely different manner by considering only the volume and the shape of the cavities -- in this case, only of the buccal cavity, since the pharyngeal cavity is fixed. This gives a more direct relationship with the notes of the formants by which we recognize the vowels -- in this case, only the note of the second formant, the one which is principally dependent on the buccal cavity. One can consider that the buccal cavity is the same size for the four nasal vowels.[6] There remain then only two variables: the degrees of opening in the front of the buccal cavity and in the back of it, which can be assigned as follows:

	BACK OPENING	FRONT OPENING	NOTE OF RESONANCE
ɛ̃	wide	wide	high
œ̃	wide	narrow	mid-high
ɑ̃	narrow	wide	mid-low
ɔ̃	narrow	narrow	low

But then we must recognize that the traits are relative: a wide front opening for /ɛ̃/ cannot be compared with a wide front opening for /ɑ̃/, but with a narrow front opening for /œ̃/; and a narrow front opening for /œ̃/ cannot be compared with a narrow front opening for /ɔ̃/, but with a wide front opening for /ɛ̃/. Whichever of the two systems we use, that of (a) or (b), we must suppose that the distinctive features only act by opposition to a different feature and not by comparison with a similar feature. Thus, the fact that the labial rounding of /ɔ̃/ is more pronounced than that of /œ̃/ does not have a direct effect, but can only be considered as a secondary effect of the backing of the tongue.

THE ACOUSTIC ASPECT (PHYSICAL)

We recall that the color of every vowel is characterized by two formants, that is, two groups of musical notes, or two frequencies (pitches) which reflect approximately the resonance notes of the pharyngeal and buccal cavities. The frequencies of the formants of all the oral vowels of French were published in The French Review sixteen years ago! (FR, 21: 477-485, 1948). One will find there for example that /i/ is characterized by a low formant and a high formant, /u/ by two low formants, and /a/ by two formants of medium pitch.

The nasal vowels do not differ in this aspect -- they are also characterized by two formants. Thus only two questions come up about them: a. How can they be distinguished acoustically from each other? b. How can they be distinguished acoustically from oral vowels? In other words, what is the acoustic trait of vowel nasality?

a. Figure 2, which is an acoustic replica of the x-rays in Figure 1, answers the first question. Nasal vowels are distinguished from each other by the frequency of the second formant, the first formants being about equal. This was to be expected since, as we have seen in the preceding pages, only the buccal cavities are different from each other, while the pharyngeal cavities of all nasals are about the same. The four nasals have the same relation to each other as the three close vowels /i/, /y/, /u/, which all have the same first formant and are distinguished only by the frequency of the second formant, as is shown in the third column of Figure 2.

b. As for what distinguishes nasal vowels from oral ones, we must look for this not in the frequency of the formants but in their INTENSITY. Whereas with the oral vowels the first formant has an intensity at least equal to that of the second, with the nasal vowels, the first is much weaker than the second. An artificial nasal vowel can by synthesized by starting out with an oral vowel with two formants and gradually reducing the intensity of the first formant. When the decrease reaches 10 to 15 decibels, the vowel takes on the nasal color that every Frenchman recognizes. Figure 2 is a schematic representation which shows, by reducing the width of the first formant, that its intensity is weaker than that of the second formant.

In addition, a low note of about 250 cycles appears in nasal vowels. This note is always present on the spectograms of nasal vowels, but it is not essential to the characterization of the nasality of the vowels. Experiments in synthesis have shown that it only contributes to the "natural" quality of the nasal vowel. (On the other hand it is indispensable for characterizing the nasality of consonants.)

In short, what we call subjectively "nasality" is simply a lack of balance between the intensities of the two characterizing formants, caused by a reduction in the intensity of the first formant. The nasal vowel can thus be described as a chord whose higher notes dominate its lower notes.

c. But the differences of frequency between the formants of nasal vowels and those of oral vowels are not entirely insignificant. They certainly contribute a little toward distinguishing nasals from orals. Figure 2 shows that these differences are small for fronted vowels: /ɛ̃/ and /œ̃/ have their first formant a little higher than /ɛ/ and /œ/. But they are large for the back vowels: the two formants of /ɑ̃/ are lower than those of /a/ -- they would also be lower than those of /ɑ/, and for /ɔ̃/ the first formant calls to mind that of /ɔ/, and the second formant that of /o/.

d. Finally the nasal vowels are generally longer under the same conditions than the oral vowels, and that should also add something to the oral/nasal distinction. In a closed syllable, the difference is noticeable: the vowels in *cinq* and *monde* are on an average markedly longer than those of *sec* and *mode*. In other positions the difference is less pronounced: the vowels of *peint* and *dompté* are slightly longer, on the average, than those of *paix* and *doté*.

THE COMPARATIVE ASPECT

Just because French has four nasal vowels out of sixteen does not mean that it can rightly be called a nasal language. On the contrary. Since nasality is distinctive, a Frenchman is forced to give a "very oral" color (timbre) to the twelve vowels that are not nasals, in order to avoid all possible confusion between orals and nasals. Words like *baie*, *fête*, must be free of all nasality so as not to be confused with *bain*, *feinte*. Aside from the four nasal vowels (which take up less than 15 per cent of the language) French gives forth a purely oral sound. On the other hand, it is languages in which nasality is not functional that would have a tendency to nasalize vowels indiscriminately. American English gives us a good example of this. Since in nasalizing all the vowels of *John came home* [dʒɑ̃n kẽim hõum] an American does not make the sentence undergo any change in meaning, there is no functional reason which prevents him from nasalizing.

In order to study the phenomenon of free nasalization as compared to functional nasalization, we made motion picture x-rays of three Americans from the West saying the sentences:

Sam in the cab saw that Ann was sad

At the camp the pond sank

and three French persons from the North of France saying the corresponding sentences:

On blâme l'Arabe de ce que l'âne est malade

La bande campe jusqu'au cinq

The results of the free nasalization can be seen in Figure 3, columns 1, 2 and 3. The frames chosen are typical of what the three Americans do.

First column. It is when the oral vowel comes before a nasal consonant followed by another consonant that the anticipation of nasality is the most marked. In *sank* [sæŋk] the velic passage already begins to open in the second frame. During the 4th, 5th and 6th frames, the soft palate (velum) is as low as it is for the French nasal vowels. In frame 7, the velum rises again and folds. In frame 8 it is in the folded position typical of nasal consonants. This makes the velic passage narrow enough to let a certain air pressure build up behind the buccal obstruction, and at the same time allows the communication which is necessary for consonant nasality between the velic and pharyngeal cavities. Finally in frame 9, the velic passage closes completely for the [k].

Second column. In *Sam* [sæm], the velic passage does not open until frame 3. During frames 5, 6 and 7, the velum is lowered in an arched position as for a French nasal vowel. In frame 8, it folds up. In frame 9 it is folded in the position typical of nasal consonants.

Third column. In *shack*, the final consonant is no longer a nasal, and yet from the 3rd frame on, the velum draws away from the Passavant ridge (small muscular bulge at the very top of the pharyngeal wall) and stays parted from it until frame 7. But during the pronounciation of the vowel, the velic passage is smaller than before a nasal consonant, as in *sank* and *Sam* -- the velum only folds up, without lowering in an arched position. In short, it is only when air pressure must build up in the mouth to prepare for the opening of a closed oral consonant that the velum of a typical American speaker comes to press firmly against the Passavant ridge and actually closes the velic passage.

Fourth column. The French word *blâme* shows how brief and weak anticipation of nasality is with a typical French speaker. However, this anticipation is not completely absent, since the velum clearly opens in frame 8, well before the lips come together for /m/.

Let us note here that with all the French speakers we filmed by motion picture x-ray, the velum lowered in anticipation of nasal consonants, but was delayed in lowering when forming nasal vowels. That is understandable since the velum cannot begin to open before the pressure of inner air for the preceding oral consonant has ceased.

THE STATISTICAL ASPECT

French nasal vowels are to be found in all positions -- initial: *ainsi*, *humble*, *anse*, *onde*; medial: *quinze*, *lundi*, *danse*, *tombe*; and final: *pain*, *brun*, *cent*, *don*. In spite of that, they come fairly low on the frequency of occurence scale: of the sixteen French vowels, /ã/ is fifth, /õ/ tenth, /ɛ̃/ twelfth and /œ̃/ fifteenth, squeezed out of last place only by the back /ɑ/. Their frequencies of occurence are 7.59 per cent for /ã/, 3.85 per cent for /õ/, 2.45 per cent for /ɛ̃/ and 1.05 per cent for /œ̃/, using equal proportions of both narrative and conversational texts as a basis. Total for the four vowels: 14.94 per cent, well under the norm which would be 25 per cent (for four sounds out of sixteen).

THE HISTORICAL ASPECT

At the time of the Roman invasions, Latin did not have any nasal vowels. It was the Gauls who, learning Latin without losing their lax articulatory habits (which made them dipthongize vowels and palatalize consonants), nasalized oral vowels every time they preceded a nasal consonant. By anticipating an /n/ or an /m/, their soft palate lowered too soon and the vowel took on a nasal color, just as our Anglo-Saxon students do today when they pronounce *bonnet* like *bon nez*. But nasalization was only the first stage of a long change -- once nasalized, the vowel changed its articulatory position. In this way *vinum* first became [vĩn], then the vowel opened to [vẽn], then to [vɛ̃n], and finally, at the very end of the Middle Ages, affected by the falling of final consonants and open syllabication which separated a vowel from the following consonant, the nasal consonant disappeared, and the pure nasal vowel of [vɛ̃] appeared such as we know it today. (At the time that plenu was turning into [plɛ̃], open syllabication was having an opposite effect on plena, which was losing its nasality and going from [plɛ̃nə] to [plɛ-nə]. This is the phenomenon of de-nasalization.) All the nasalized vowels changed color in this way: just as [ĩn] became [ɛ̃], [ỹn] became [œ̃], [bõn] became [bɔ̃], [ãn] became [ɑ̃], with the articulatory characteristics shown in Figure 1. Apparently it was under the influence of nasality that these color changes came about, because all nasal vowels finally reached a point at which they had a phonetic characteristic in common which they share with no other vowel -- today all French nasals have the same pharyngeal cavity and the same first formant.

But French nasality was not always distinctive. At first it was undoubtedly non-functional, as in English, where you can pronounce [mæ̃n] as well as [mæn] and nothing other than man is understood. It was only from the tenth century on that nasality began to play a linguistic role. In the Chanson de Roland, /a/ no longer rhymes with /ã/, which leads us to believe that /ã/ was already pronounced [ɑ̃] or at least something clearly different from [ã]. Thus we must assume that the frequency of the first formant, shared by all the nasal vowels, and reflecting a given volume of pharyngeal cavity common to all of them, represents the optimum condition for nasalization and so facilitates distinctive nasality.

After the Renaissance, nasal vowels seem to have continued their evolution in the North-West of France. One finds today that Brittany and Ile de France have partially lost the vowel /œ̃/, which has assimilated into /ɛ̃/. There is an explanation for this state of events. a. The frequency of occurence of /œ̃/ is very low: only 1.05 per cent of the vowels, in spite of the article un. b. Owing to the weakness of the first formant, which we discussed earlier, the color of nasal vowels is not so well defined as that of the oral vowels. As it happens, their evolution has brought them to the area of the phonetic triangle where the vowels are nearest to each other. So it was to be expected that the front nasals on one hand and the back nasals on the other hand become identical. c. If the merging is further along for the front nasals than for the back nasals, it is perhaps due to their frequencies of occurence which are much greater for the back nasals: 7.59 per cent and 3.85 per cent, total 11.64 per cent for /ɑ̃/ and /ɔ̃/, and only 2.45 per cent and 1.05 per cent, total 3.50 per cent for /ɛ̃/ and /œ̃/.

THE THEORETICAL ASPECT

Now let us look for the physiological source of nasality.

The data on which to base a theory are the following: a. The historic facts explain the articulatory facts: once the vowel was slightly nasalized by anticipation of a consonant, it evolved until its pharyngeal cavity reached an optimum size which is the same for all nasal vowels. b. It is upon reaching this optimum dimension that the first formant was weak enough to produce a distinctive nasality.

Why then is the first formant weak? Let us go back to the x-rays of Figure 1. Just above the soft palate is a small cavity, called velic cavity, the only nasal cavity that has walls smooth enough to facilitate resonance -- the rest of the inner nose is covered with fibrous, absorbant walls. When the velic passage is open, the pharyngeal and velic cavities are connected and resonate harmonics (produced by the vocal cords) which coincide with the natural resonance notes of these cavities. Now, when the velum is lowered, these two cavities have about the same dimensions and resonate the same harmonics. If a harmonic resonated in the pharyngeal cavity is at the same time resonated in inverse relation in the velic cavity, the pharyngeal resonance of this harmonic is dampened or cancelled. Not only one, but several harmonics of the first formant can be weakened or cancelled in this way. It is, therefore, by adjusting the volume of the pharyngeal cavity (variable) to that of the velic cavity (fixed) that the most favorable conditions for the weakening of the first formant are created, that is, the conditions which are most favorable to vowel nasality.

THE PEDAGOGICAL ASPECT

Here four different problems arise that we must consider separately: distinction of color between nasals and orals; distinction of color among the nasals; consonantal anticipation at the juncture of a nasal vowel with an oral consonant; and consonantal anticipation at the juncture of an oral vowel with a nasal consonant. In this last case the difficulty encountered is not to nasalize, but to give the vowel a clearly oral quality, that is, to produce vowels without any trace of nasality, even before a nasal consonant.

a. Foreign students have no trouble nasalizing vowels. Anglo-Saxons find it even too easy, and their real problem is to keep from nasalizing oral vowels, as we shall see in (c).

b. They have more trouble when it comes to producing the exact color, to making a clear distinction among the four colors. Two reasons are attributed to this difficulty. The first is that, since nasal vowels are characterized almost only by the second formant (we have seen that all the first formants have the same frequency and are very weak), their color is more vague, less well differentiated than that of the oral vowels. The second reason is that nasal vowels do not "correspond" to oral vowels, especially as concerns /ɑ̃/ and /ɔ̃/, as the x-rays of Figure 1 and the spectograms of Figure 2 clearly indicate. Thus what the student has to learn is not four modifications of oral vowels, but four new vowels. Those who teach the nasal vowels by starting out with an oral vowel that is then nasalized are making a serious error. The nasals must be taught directly, without going by the so-called corresponding orals. The most effective way is to introduce them to the student orally (without spelling) in words, or better yet in a sentence, after the student has mastered the color of oral vowels. Pedagogic experiments have shown that, under normal language-class conditions followed by a little work in the laboratory, a sentence of nasal vowels must be worked on audio-lingually for three to five days in a row before the auditory and articulatory habits can be established with accuracy. This may be a long time for four vowels, but it is the only way to do the task right from the very start so that it will never have to be repeated. After that, the words of the sentence can be used as keys for all other nasals, and correction from the professor is no longer necessary.

But the real teaching problems do not stem from vowel color but from the American habit of letting nasality spread.

c. When a nasal vowel is followed by an oral consonant as in onde [ɔ̃d], the least bit of consonantal anticipation causes a nasal homorganic consonant (having the same point of articulation) to be inserted: an [n] is inserted before /t,d,s,z /, and honte, onde, once, onze become [ɔ̃nt, ɔ̃nd, ɔ̃ns, ɔ̃nz]; an [m] is inserted before /p,b,f,v/, and pomper, bomber, confie, convie become [pɔ̃mpe, bɔ̃mbe, kɔ̃mfi, kɔ̃mvi]; an [ŋ] is inserted before /k,g/, and tank, tangue, become [tɑ̃ŋk, tɑ̃ŋg].

In order to become conscious of the mistake and of the fact that it is caused by the psychological phenomenon of consonantal anticipation, the student must listen to himself saying a series of reversible words like béton, béton, béton for tomber, tomber, tomber, and he will realize that he does not put in an extra consonant in the first series as he does in the second.

The best corrective exercise consist of making the following kind of distinctions so as to be understood perfectly by a Frenchman:

Au tome B	[o tom be]
Haut tombé	[o tɔ̃-be]

C'est comme blé	[se kɔm ble]
C'est comblé	[se kɔ̃-ble]
Quelle bonne théière	[kɛl bɔn tejɛr]
Quelle bonté hier	[kɛl bɔ̃-te jɛr]
On a scène D	[ɔ̃ na sɛn de]
On a scindé	[ɔ̃ na sɛ̃-de]

d. When an oral vowel is followed by a nasal consonant, as in Jeanne, Rome, consonantal anticipation causes the greater part of the oral vowel to be nasalized, as is convincingly shown on the radiograms of Figure 3. To make the student conscious of the mistake and of the anticipation which causes it, the professor should have him compare reversible series like Jeannet, Jeannet, Jeannet, and neigea, neigea, neigea, and help him to hear that he nasalizes the /a/ in the first series and not in the second series.

The best corrective exercise is again based on practicing the pronunciation of minimal pairs. The student should manage to make the following kind of distinctions and be understood perfectly by a Frenchman:

Un bon nez	[œ̃ bɔ̃ ne]
Un bonnet	[œ̃ bɔ-ne]
J'en ai là	[ʒɑ̃ ne la]
Jeanne est là	[ʒɑ-ne la]
Mon automne	[mɔ̃ notɔn]
Monotone	[mɔ-nɔtɔn]
Un certain nerf	[œ̃ sɛrtɛ̃ nɛr]
Un certain air	[œ̃ sɛrtɛ-nɛr]
Pas de bon nombre	[pa d bɔ̃ nɔ̃br]
Pas de bonne ombre	[pa d bɔ-nɔ̃br]

THE AESTHETIC ASPECT

Poets seek out nasal vowels for what they generally call their softness. When asked to be specific about the impression the vowels produce, writers also use such adjectives as veiled, obscure, indirect, nebulous, shaded, grave, majestic, sensuous. These subjective qualifiers correspond rather well to the objective facts that we introduced above when speaking of the acoustic composition of nasal vowels: a weakness of intensity in the low notes as compared to the high notes inevitably makes the color more vague. Experiments with synthesis have shown that the more the intensity of the low notes is reduced, the more indefinite the vowel color becomes. But apart from the lack of balance of the intensities, which in itself is responsible for what is called nasality, we must take into account the total intensity, which is weaker, and the holding, which is longer in the nasals than in the orals, all else being equal. These two characteristics would explain the impression of solemnity and majesty which go along with haziness. One can judge for himself by examples in which nasal sounds dominate. In the following lines from Baudelaire's Correspondances, the proportion of nasals is 25 per cent, whereas in ordinary prose it is less than 15 per cent.

"Comme de longs échos qui de loin se confondent,

....

"Les parfums, les couleurs, et les sons se répondent.

....

"Comme l'ambre, le musc, le benjoin et l'encens,
"Qui chantent les transports de l'esprit et des sens."

To close this study, let us quote from a description of nasal vowels a Japanese engineer was prompted to write while doing an electronic analysis of them in his laboratory. It offers a moving account of the poetic insight of the Japanese, and at the same time an unprejudiced subjective judgement.

"There is no language which does not contain nasal inflexions in its consonantal system, but there are very few that have true nasal vowels. French can be counted among the latter. The color of French nasal vowels seems to us to be of unparalleled charm. The nuances characteristic of French speech are partly due to the frequency occurence of rich nasal sonorities. It is fascinating when, for example, vaporous nasals are uttered by a rich low-contralto voice. In the course of a French conversation, and especially during the reciting of poetry, the nasal vowels stand out in subdued relief against the rhythmic flow of sound waves. This mood of beauty would suggest that the modulated vowels of nasality form an exquisite contrast with the oral vowels -- a contrast which puts the listener under a spell resulting from the alternation of oral and nasal timbres, of bright, gay, lustrous colors and pale, indecisive, nebulous ones. This elucidation takes us to the very heart of the distinction between orality and nasality, in the very essence of their nature. In figurative terms, the colors of orals converge in a bright sunny image, and those of nasals spread shaded pools of light."[7]

Footnotes

VOWEL NASALITY IN FRENCH AND ENGLISH

[1] The research reported herein was performed pursuant to a contract with the U.S. Office of Education, Department of Health, Education and Welfare.

[2] The term "velic" refers to the end of the soft palate, or velum, which moves to open and close the velic passage by which the nasal cavities can connect with the buccal cavities or be separated from them. For, let us make it clear, the fact that the expired breath is able to "pass through the nose" is not what produces nasal color -- the distinction passe/pense is made just as clearly with the nose obstructed as with the nose open. What counts is that the resonance of the nasal cavities combines with that of the buccal cavities. Thus the combined resonance, which goes by the name of "nasal" vowel, is emitted by the buccal orifice and not by the nasal orifice.

[3] Let us note in passing that this denasalization is what the first phoneticians, about 1885, under the direction of Paul Passy, used as a basis to establish the phonetic symbols /ɛ̃/, /œ̃/, /ɔ̃/. For /ɑ̃/ they prefered the back /ɑ/ rather than the middle /a/, making the somewhat summary judgement that the tongue position of /ɑ̃/ was very far from that of /a/. (Unfortunately it is also quite far from that of /ɑ/!) If they had been able to see x-rays of the nasals, they might have suggested new symbols having no connection with the oral vowels.

[4] If we start by talking about cavities it is because, in the final analysis, the volume and the aperture of the cavities is what determines the frequency of the formants, and consequently what characterizes vowel color. The opening of the jaws, the height and fronting of the tongue and the rounding of the lips are only secondary factors, and in contributing to the shape of the cavities, they can be subjected to appreciable compensations.

Footnotes

VOWEL NASALITY IN FRENCH AND ENGLISH

[5] The pharyngeal cavities of /ɛ̃/ or /œ̃/ are actually larger than those of /ɑ̃/ or /ɔ̃/, but their form being more tubular than spherical gives them a note that is not so low as it would be for the same spherical volume.

[6] The four nasal vowels are to each other like the three closed vowels /i/, /y/, /u/ (third column) which have about the same size pharyngeal cavity (very large with very small opening) and which are only distinguishable from each other by the front cavity, which is small and open in front for /i/ (very high note), small and close for /y/ (less high note), large and close for /u/ (very low note).

[7] Yochizuki Ochiai, "Memoirs on nasalics," Memoirs of the Faculty of Engineering, Nagoza University, 9:147-153, May 1957.

Figure Legends

VOWEL NASALITY IN FRENCH AND ENGLISH

Figure 1. Profiles of nasal and oral articulatory POSITIONS drawn from motion picture x-ray films taken at 24 frames per second. On these profiles are shown the tongue position, jaw opening, the degree of lowering of the soft palate, and especially the shape and size of the buccal, pharyngeal and velic (above the velum) cavities. First column: the nasal vowels of *un bon pain blanc*, in order of decreasing maxillary opening from top to bottom: *pain*, *un*, *blanc*, *bon*. Second column, top to bottom: the oral vowels of *père*, *peur*, *part*, *port*. Third column, from top to bottom: the oral vowels of *dit*, *du*, *doux*, *dos*.

Figure 2. Acoustic representation of the nasal and oral vowels of Figure 1, by the first two formants, as shown on spectograms. The frequencies of the formants are given in cps (cycles per second). The four nasals (first column) are distinguished from each other only by the frequency of the second formant, and are distinguished from the oral vowels (second and third columns) by the weakness of the first formant (narrow line at 550 cps) and by the addition of a weak tone (very narrow line at about 250 cps).

Figure 3. Profiles of articulatory MOVEMENTS (9 successive frames) drawn from motion picture x-ray films taken at 24 frames per second. Here nasalization can be observed by the lowering of the soft palate. First column: *sank*, as said by an American; the soft palate opens as early as the second frame. Second column: *Sam* as said by an American; the velum opens in the third frame. Third column: *shack*, as said by an American; the velum opens as early as the third frame but it is never as open as in *sank* and *Sam*. Fourth column: *blâme*, as said by a French person: the velum opens only in the next-to-last frame.

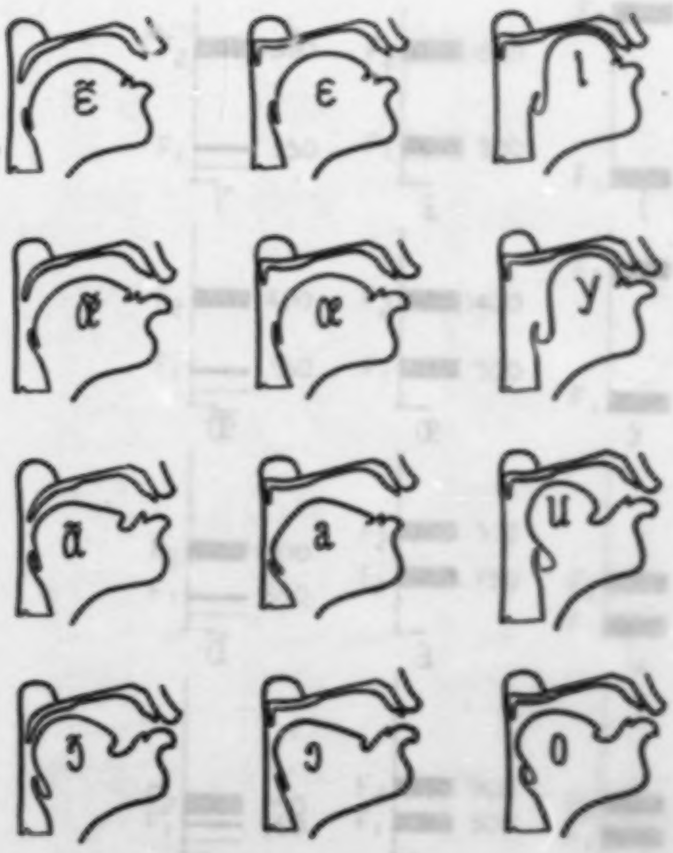

ɛ̃
ɛ
i
œ̃
œ
y
ɑ̃
a
u
ɔ̃
ɔ
o

Figure 1

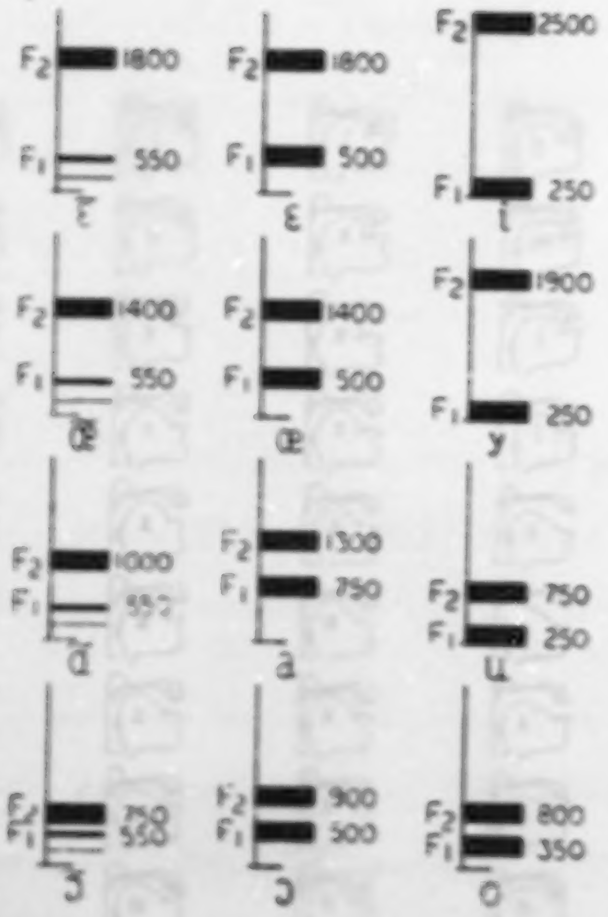

F2 1800
F1 550
ɛ̃
F2 1800
F1 500
ɛ
F2 2500
F1 250
i
F2 1400
F1 550
œ̃
F2 1400
F1 500
œ
F2 1900
F1 250
y
F2 1000
F1 550
ɑ̃
F2 1300
F1 750
a
F2 750
F1 250
u
F2 750
F1 550
ɔ̃
F2 900
F1 500
ɔ
F2 800
F1 350
o

Figure 2

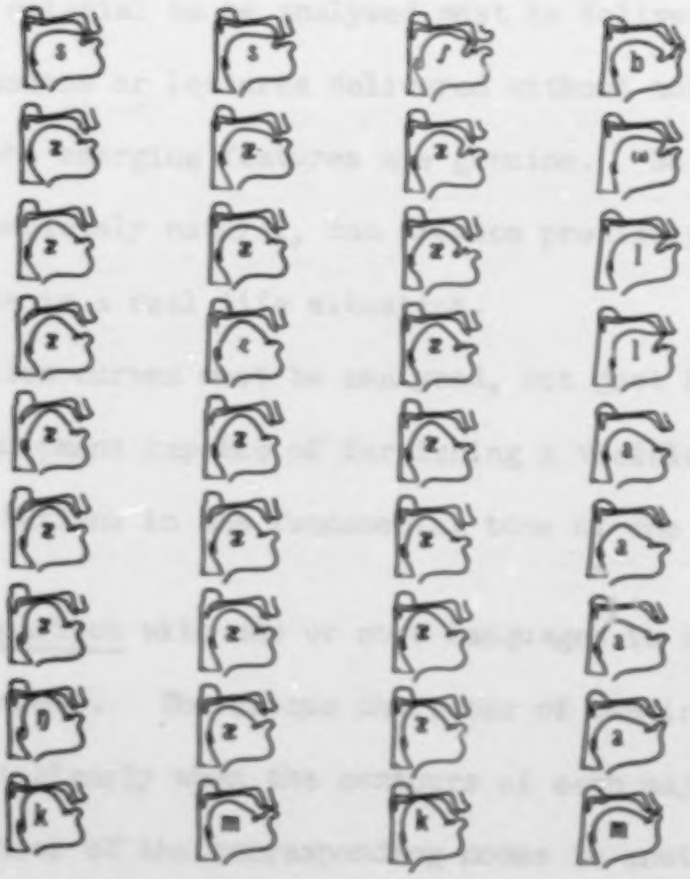

s
s
ʃ
b
z
z
z
z
z
l
z
c
z
l
z
z
z
a
z
z
z
a
z
z
z
a
ŋ
z
z
a
k
m
k
m

Figure 3

APPENDIX V

SOME CHARACTERISTICS OF GERMAN INTONATION FOR THE EXPRESSION OF CONTINUATION AND FINALITY[1]

The value of an investigation depends on its research technique. In order to study very objectively the characteristic intonation contours of a language, no less than three conditions should be observed:

a) One must look for the intonation features that all the normal speakers of a given dialect have in common. Practically speaking, the investigator must compare a good number of speakers, until the features they share emerge strikingly.

b) The speech material to be analyzed must be delivered spontaneously, as in interviews, discussions or lectures delivered without notes. Only then can one be assured that the emerging features are genuine. No reader, even the best actor, can be completely natural, can produce precisely enough the intonation that he would use in a real life situation.

c) The intonation curves must be analyzed, not just by ear, but with the help of electronic equipment capable of furnishing a visible graphic representation of frequency variations in the fundamental tone of the voice.

Furthermore, comparison with one or more languages is a great help in visualizing the typical aspect. The unique character of the intonation of a language will appear most clearly when the contours of each major mode of expression are contrasted with those of the corresponding modes in another language -- when the manner in which one language distinguishes two different modes of expression (question vs., command, for example) is compared with the corresponding distinction in another language.

As has been suggested in the preceding paragraphs, the intonation of a language may be considered under two aspects, one distinctive, the other non-distinctive. The distinctive aspect is concerned with the contour differences that distinguish one mode of expression from another within a given language, for instance, interrogation question: are you ready? from information question: who is ready? The non-distinctive aspect is concerned with the intonation shapes which, being found in many modes of expression without playing an essentially distinctive role, recur so frequently that they help to distinguish one language from another, at a distance, when words themselves may not provide a clear cue. The latter aspect, more elusive than the distinctive one, is perhaps more important in teaching; for it should permit a student to acquire a basic habit of the target language very early, with the help of only one or two simple notions.

Ambitious as it may look, both aspects will be investigated here. But in order to obtain meaningful results, we have limited ourselves to three modes of expression found in declarative intonation. These modes will be called minor continuation (A), major continuation (B), and finality (C), as in:

The characteristic (A) of his intonation (B) is a rising curve (C)

We shall also inevitably meet a few cases of word emphasis.

Let us illustrate more clearly what distinguishes minor continuation from major continuation. In the sentence John left Henry running fast to find out who had come, one can make it clear that Henry ran, not John, by using an inflection of minor continuation on Henry and of major continuation on fast:

John left Henry (A) running fast (B) to find out who had come (C)

If the inflection of major continuation is shifted to Henry and that of minor continuation to fast, it becomes clear that John ran, not Henry:

John left Henry (B) running fast (A) to find out who had come (C)

The length of the pauses after *Henry* or *fast* also contributes to this distinction of meaning, naturally, but we are interested here only in the role that pitch plays.

Although the major continuation inflection (B) may end an initial phrase, as in John left Henry (B)..., its main function is to unite several minor sense groups into a major one, as in:

The rising curve (A), the falling curve (A) and other features (A) of his intonation (B) are very distinctive (C)

THE DISPLAY OF INTONATION

During the past few years, we have built a library of tape recordings of spontaneous speech by German, Spanish, French and American speakers, and we have made spectrograms at two different scales of 5 minutes of running speech for each subject. These spectrograms make possible the measurement of many phonetic features -- among them the measurement of frequency variations as they are reproduced in Figures 1 to 4. One of the two scales, at 2000 cps per inch, essentially shows the formants and therefore the boundaries of vowels and consonants, but it is also used to observe frequency variations by following one of the higher harmonics -- in this investigation the tenth harmonic was selected throughout. Occasionally, though, the tenth harmonic may disappear; that is why another scale of the spectrographic display is useful. The second scale, at 200 cps per inch, amplifies by ten times the frequency variations of the lowest harmonics so that the deviations of the first harmonic (fundamental frequency) appear ten times as large as on the 2000 cps scale and therefore equal to the deviations of the tenth harmonic of that scale. Between the two scales, one confirming the other, the variations of frequency appear clearly enough to make their transfer possible to graphs such as those of our figures.

It is assumed here that the subjective impression of pitch variations (intonation) depends mainly upon the objective frequency variations that are displayed on the spectrograms. (The possible contribution of intensity, duration and color to pitch variations has not yet been determined and will, in the meantime, be ignored.) On our graphic display (Figures 1 to 4), the relative time scale (about four seconds per line) is proportional to reality during the voiced portions of the display (mainly vowels, liquids and nasals); the voiceless portions (mainly voiceless consonants) and the pauses are drastically reduced in order to save space. (Pauses, mainly hesitations, occupy considerable time in spontaneous speech and, curiously enough, they seldom coincide with the punctuation -- in fact, they generally occur soon after the beginning and well before the end of a sentence. Most frequently the end of one thought and the beginning of the next unite without interruption. It will also be noted that in spontaneous speech, speakers often use a continuation inflection where one would expect a period, or a finality inflection where one would expect a comma. This is the price an investigator must pay in order to obtain genuine naturalness.)

In the figures, the ends of sense groups are marked A for minor continuation, B for major continuation, and C for finality. The mark E indicates emphasis on a word. All four marks were determined by ear, through comparison and contrast, by a group of linguists; the pitch graphs indicate the shape of their acoustic realization. The A, B and C marks are placed under the last unstressed syllable of a sense group (or the end of a stressed syllable if no unstressed ones follow), because it is the unstressed end of a sense group that is most critical in distinguishing B from A, or either one from C. The mark E, on the contrary, is placed under the stressed syllable of the word upon which emphasis falls.

THE RECORDED MATERIAL

Four representative passages of spontaneous speech by cultivated speakers of North German were selected from our collection of tape recordings and spectrograms for Figures 1 and 2. We shall look for what these four subjects have in common in regard to intonation habits.

Since an American teacher of German must not only know German intonation but also the intonation habits of his own tongue, including those of his students' dialects, Figure 3 presents for comparison a passage of typical American intonation by a cultivated speaker. It is an example of spontaneous speech, taken from a series of lectures delivered entirely without notes. Finally, Figure 4 presents two selections from Tonio Kröger read by Thomas Mann himself. This will offer an opportunity to compare the diction of prepared reading to that of improvised conversation. Naturally, our statistics do not include passages which are read, but only freely spoken ones.

The first German speaker (Figure 1) is a German girl (K) who was asked to tell the story of her flight from Germany during the war years. She sat before the microphone and spoke without preparation. The first passage that was selected from speaker K's tape illustrates four series of continuation sense groups; the second, two finality sense groups. The other German subjects are males. The second speaker (Figure 1) is the president of the Volkswagen Company (L). In a radio interview, he answers questions about his ascent to this high post after the war (EMC Corp. Tape No. DTG 1000). He illustrates without interruption three series of continuation groups and two oi finality. The third speaker (Figure 2) is the interviewer (M) of the Volkswagen Company president. In the selection we present, he illustrates two pairs of continuation-plus-finality series that are exceptionally well balanced for spontaneous speech. The fourth speaker (also Figure 2) is an eminent medical doctor and professor (N) who is being interviewed about his research. This uninterrupted passage is spoken with considerable emotion. It was selected to illustrate word emphasis and asymmetry in the sequences of A, B and C inflections, while the characteristic intonation shapes are retained. The first sentence starts with a B inflection. The next has neither A nor B, in spite of its length -- 23 syllables uttered in a single flow. The six lines contain as many as five word emphases, which cause considerable disturbance in the inflection patterns.

The range of frequency variations is indicated by the horizontal lines which are drawn at the lower limit of the mode of finality and at the upper limit of the mode of major continuation. Contours of emphasis will often reach higher than those of major continuation.

The range of K is barely one octave (6 tones). She speaks in a rather bored and monotonous voice. But in spite of this lack of interest, she shows the same general characteristics as the enthusiastic voice of N with a range of 10 tones. The range of L is a little less than an octave and a half (8 tones); that of M a little more than an octave and a half (9.5 tones); and that of N, 7.5 tones without emphasis and 10 tones counting his exceptional insistence on nur. Briefly, the range of the three men approximates an octave and a half. We shall see later that the reading diction of Thomas Mann carries him to a range of more than two octaves (14 tones).

We can now proceed to an acoustic description of each German type of contour, to be followed by the manner in which these contours contrast among themselves. Then we shall compare German with English, and finally, reading diction with improvisation.

Our four recordings of extemporaneous speech offer altogether 36 A (minor continuation), 13 B (major continuation), 10 C (finality), and 8 E (word emphasis) contours.

THE SHAPE OF CONTINUATION

Let us find out first what pitch features are common to both modes of continuation -- minor continuation (A) and major continuation (B). We shall refer to specific examples in the four texts of Figures 1 and 2, as well as to Figures 5 and 6, in which all German examples are assembled.

A concrete image -- that of a bird -- will simplify our description (see lower right hand corner of Figure 1). The complete pitch pattern of continuation, at the end of a German sense group, invariably takes a shape which suggests the body outline of a bird singing, seen from profile with its tail on the left and its beak raised up on the right. The bird can be nearly outlined, as in geboren (Kb-2), Verbannung (La-6), or merely suggested, as in dort (Ka-4), kennen (Mb-1). The completion of its shape seems to depend on the degree of stress involved. Having the bird picture in mind, it will be convenient to divide the pattern into a tail (dip, falling pitch), a back (depression, fall-and-rise pitch), a neck (rising pitch in tilted S-shape), a head (high pitch plateau), and a beak (prolongation of high plateau for unstressed syllables) -- a duck's beak is perhaps the most appropriate image here.

The back. Every continuation pattern shows a depression in the back of the bird shape (see Figure 5), so that from back to neck there is always at least a suggestion of the lower half of an S-shape. This back depression before the rise of the neck is the most characteristically German section of the bird shape, as we shall see later. It contains the tune by which German is recognized at a distance.

The tail. A clear pitch-fall, or dip, occurs as the tail of the bird shape in 41 of the 49 A or B patterns (see Figure 5); and in the remaining 8 there is barely more than a suggestion of a tail. The more the tail is defined, naturally, the more pitch depression of the back -- which is so characteristically German -- is noticeable. The tail is therefore, like the back, a very German feature of pitch.

Note that the dip in the tail does not always begin in the stressed syllable, as in Glauert (Kb-1), fragte (Lb-3). It can perfectly well start in the unstressed syllable that precedes, as in gelebt (Kb-3), or it can even involve -- but very exceptionally -- more than one preceding unstressed syllable, as in Generaldirektor (Ma-9). Of the 49 A or B contours of our four subjects K, L, M, N, 20 tails seem to start on an unstressed syllable and 29 are entirely in the stressed syllable.

The neck. Except in two cases (out of 49), the occurrence of maximal stress is in the back and the neck. This is mainly what justifies one in calling German continuation "rising." (In English, as we shall see later, the occurrence of maximal stress is in a falling portion.) The two exceptions found here are kénnen (Mb-1) and Génius (Ma-6), which have the stressed portion in the falling tail and the unstressed in the rising neck. (Nevertheless, these two words retain the bird pattern.) It can also happen that a secondary stress takes the rising-neck shape away from the primary stress. We have a single case of it in Militärregierung (La-8), where the primary stress is falling and the secondary stress rising. And there are long or compound words in which both stresses take a bird shape, although less pronounced in the secondary than in the primary, as in Volkswagenwerke (Ma-10), Tschechoslovakei (Ka-7), hängenblieben (Kb-4).

The head. In our four selections there are 11 continuation contours in which the stressed syllable is final (column 1 of Figure 5), 28 in which one syllable follows (column 2), 4 in which 2 syllables follow (column 3), 4 in which 3 syllables follow (column 4), and 2 in which 4 syllables follow (column 5). This offers a good range of different occurrences for investigation.

Examples of each type are:

0	Flúcht	(Ka-6)
1	Léitung	(La-9)
2	wúrde ich	(La-2)
3	hä́ngenblieben	(Kb-4)
4	Vólkswagenwerke	(Ma-10)

In all but 4 of the 49 continuation contours, the head (the top of the neck-rise) is higher than the back or the tail. In four exceptions the head rises only to the level of the tail, as in the weak stressing of Bréslau (Ka-2).

The beak. Here it represents the unstressed syllable, or syllables, which follow a stress (columns 2, 3, 4, 5 of Figure 5). The beak is the most critical portion of the bird shape in regard to expressing continuation or not expressing it. Compare hören (after Ma-3) to werden (Ma-4), Doktor (after Ma-7) to Nordhoff (Ma-8). Both hören and Doktor, with their beaks pointing down, fail to indicate continuation; whereas werden, Nordhoff, with their beaks pointing slightly up, hold a clear cue to continuation. Compare also Wolfsburg (Mb-2) with Europas (Mc-2). After similar back-and-neck shapes, ...burg cues to continuation with its rising beak, ...pas cues to finality with its falling beak.

We shall consider later the patterns which involve word emphasis in addition to continuation (marked e in Figure 5). With one syllable following the stress (column 2), the beak (the unstressed syllable) is barely lower than the head in 4 cases, about the same in 5 cases, a little higher in 8 cases, and much higher in 11 cases. With 2, 3, or 4 syllables following the stress (and not involving word emphasis), the weak syllables are barely lower than the head in one case, nearly at the same level in one other case, and clearly higher in the other four cases. In short, the post-stress syllables that cue to continuation are on a high plateau at the approximate level or higher than the head.

The one exception to this generalization occurs when word emphasis is involved in the continuation pattern. Then a disturbance occurs which may send the unstressed syllables lower than normal, as in ...FÜNFundvierzig (Ke-1, a-5), ERSTer Linie (Ne-4, a-4). But, in any case, the pattern retains a cue to continuation in a slight rise at the very end, in the last unstressed syllable.

There are five other cases of word emphasis (column 6 of Figure 5), for a total of eight, in our statistics. These five cases are independent of what follows them. They start in the same manner as the ones which are involved in continuation -- a bird shape with an abnormally high head, but there is no cue to continuation in the following syllables (if any): GRÖSSte (Me-1), KENne (Ne-1), NUR (Ne-2), erKENne (Ne-3), and WIE (Ne-5).

To summarize, the only feature that the eight cases of word emphasis have is a very high head -- at least as high as the average B level -- immediately followed by a lower syllable.

A good way to realize the effect of word emphasis is to compare similar words in similar conditions, with and without emphasis. We have here the cases of Volkswagenwerke(s) in La-10 and in Ma-10. With word emphasis by L, who wants to contrast it with Opel, it has a very high head and low beak; without word emphasis, by M, it has a medium high head followed by a rising beak. Similarly, in Ka-5 neunzehnhundertFÜNFundvierzig, fünf has a very high head and a lower beak. But a few sentences later (not reproduced here) K pronounces neunzehnhundertsechsundvierzig without emphasis, giving to sechs a medium high level followed by three slowly rising syllables. These contrasts are striking, and if we compare them with others in our recordings of spontaneous speech, they are typical of German.

THE SHAPE OF FINALITY

Of the ten finality contours that are found in our four selections (see Figure 6), three are in the final stressed syllable, two have the stress followed by one unstressed syllable, three by three unstressed syllables, one by four unstressed syllables, and one by seven unstressed syllables. Examples of each type are:

0	sagt	(Nc-3)
1	Europas	(Mc-2)
3	Hochdeutsch sprechen	(Kc-1)
4	verstanden zu werden	(Kc-2)
7	Volkswagenwerke gekommen	(Lc-2)

Let us except Mc-1, in which a French-like pattern is used (after a rise on the last syllable of Genius to express continuation and the end of a sense group, the final fall of wissen begins only a little lower than the end of the preceding group), and generalize on the basis of the other nine finality patterns.

Finality contours begin with a back depression and a neck rise that are not very different from those of the B patterns. In fact, maximal intensity is on the rising neck, as it was in continuation, so that the head is high and loud in all cases. But following the head, the beak points down (when no unstressed syllable follows), or lies low and nearly flat, cut off from the head (when unstressed syllables follow the stress). In short, the profile of finality is not that of a bird singing but rather of a bird eating. But it should be described differently, depending on whether or not unstressed syllables follow the last stressed one.

a) If no unstressed syllables follow, as in Person (Nc-2), sagt (Nc-3), spricht (Nc-4) (column 1 of Figure 6), the bird characteristically has a back depression, a rising neck, a high head and a long, sharply falling beak strikingly resembling that of a stork.

b) If one or more unstressed syllables follow the stress (columns 2, 3, 4, 5 of Figure 6), as in übernehmen wollte (Lc-1), the bird characteristically has a falling tail, a back depression, a rising neck, a high head, a pitch break, and a low plateau beak lying nearly flat on the floor.

The early location of intensity on the rise, in the pattern of finality, cannot be emphasized too much, however. Where no unstressed syllables follow, maximal intensity is on the rising neck, not on the downward beak -- the beak has a very low intensity (contrary to English, which puts the maximal intensity on the falling portion). Similarly, where unstressed syllables follow the stress, the intensity of the rising neck is high and that of the unstressed syllables is low. Thus, the pitch break impression which occurs between the head and the beak (low unstressed syllables), is even more pronounced for the ear than it appears on the graphs.

Measured in tones, the finality breaks are of 3 and 2 tones by K, 2.5 and 2 tones by L, 3.5 and 2.5 tones by M, 3 tones by N, for an average of nearly 2.5 tones. The intervals between the highest point of the head and the lowest point of the beak (now including the last syllable stresses) are of 5 and 4 by K, 5.5 and 4.5 by L, 5 by M, 5, 4, 7 and 5.5 by N, for an average of nearly 5 tones.

MEANINGFUL CONTRASTS

1. Finality versus continuation. In German this contrast is clear, but it occurs very late -- only in the beak of the bird shape (compare Figure 6 with Figure 5). Just as in German syntax the key verb is often revealed at the end of a phrase, so in intonation the key portion is not known until the last pitch element has been heard. The back and neck of finality can well sound like

those of continuation -- especially major continuation; it is only by the high beak of continuation or the low beak of finality that one is finally informed of whether the sentence ends or is to continue. Most different from this is perhaps the Spanish pattern, in which a high unstressed syllable __before__ the last stress already announces the inevitable occurrence of finality:

Spanish: yo TENGO QUAtro

German: ich HAbe SIEben

2. __Minor continuation versus major continuation.__ The dip, rise and high plateau that characterize continuation are clearly more pronounced, more completely defined, in major continuation than in minor continuation (compare the upper to the lower portion of Figure 5). Using the bird image, on the average, for major continuation the back depression is deeper, the neck is longer, the head is higher, and the beak is more rising. The cue is mostly in a higher and more rising beak. The three other sections generally contribute to and announce the more rising beak, but if they fail to do so, a rise of the last unstressed syllable over the preceding one is a sufficient cue. On the other hand, the higher the head, the less the height and rise of the beak are utilized. Measured in tones (not including the three cases that involve word emphasis) the musical interval between the deepest point of the back and the highest point of the head or beak varies, for minor continuation, from 0.5 tone in __Schlesien__ (Ka-3) to 4 tones in __Opel__ (La-3); and for major continuation, from 1 tone in __kennen__ (Mb-1) to 8 tones in __entlassen__ (Lb-1). The average rise is almost 2 tones for minor continuation and just over 4 tones for major continuation. By subjects, the A and B intervals are respectively 1.66 and 2.75 tones for K, 2.1 and 5 tones for L, 2 and 3 tones for M, 2 and 5.75 tones for N.

LANGUAGE CHARACTERISTICS (PHONETIC)

The intonation features which are rather particular to German and permit one to recognize it by ear at a distance are the depression of the back and the break between head and beak. The latter has a much lower frequency of occurrence than the former, since it is limited to the contour of finality (the proportion is perhaps about 1 to 10).

1. The back depression. For speakers of Spanish, English and French, stressed syllables of German are subjectively heard as dragging. This impression can perhaps be assigned objectively to the pitch dip that precedes the rise. This dip, repeatedly observed on our intonation graphs (Figures 5 and 6), delays the rise and the moment when the stressed vowel reaches its pitch peak. Other languages may have just as long an average stressed vowel, but they reach the pitch peak faster, more directly, without the detour of the back depression. This back depression shape is not only in the continuation, finality and word emphasis patterns, it is also found in the other stressed syllables. In our four selections we count 42 of these, many of them with a complete bird shape but with beak directed downward, as in Schule (after Ka-9), automatisch (after Ka-10), so that in our two pages of graphs (200 words) the drag impression caused by a pitch dip recurs a total of 110 times, more than once every other word. One can see how much help a student of German could receive from taking this single intonation habit.

2. The beak break. After the back depression, what is most characteristic of German intonation is the sudden pitch break, from high to low, that separates the unstressed syllables from the last stressed one in finality (Figure 6). The best examples here are HOCH/deutsch sprochen (Kc-1), halluziNIE/re ja nicht (Nc-1). The sudden lowering of pitch and intensity on the last three syllables gives a foreigner the impression that deutsch sprechen, re ja nicht must be

separate words. English and French show no discontinuity comparable to German, which makes hofstrasse, in the following example, appear as a separate semantic entity.

English: windowwasher

French: salle à manger

German: bahnhofstrasse

COMPARING GERMAN AND AMERICAN ENGLISH INTONATION

To compare now German and American intonation habits, let us use a page of contour graphs (Figure 3) obtained by spectrographic analysis of the extemporaneous speech of a typical American speaker. The detailed study of frequency contours in American English was done in two previous studies for comparison with French[2] and with Spanish[3], and will not be repeated here. We shall simply compare the most characteristic features of American intonation which emerged from the previous studies with the most characteristic features of German intonation which have just emerged from the present study. Let us call the American speaker H. (Note that in Figure 3 the continuation and finality contours of the American speaker are marked in capital letters -- A, B, C -- instead of lower case -- a, b, c -- as for the German speakers.) English will always mean American English in what follows.

1. Characteristically, German stressed syllables begin with a depression (followed by a rise), English ones with a hump (followed by a fall). The

following comparisons are convincing:

weiss (Nb-2)	know (HA-1)
grösste (Me-1)	human (HA-3)
gelebt (Kb-3)	disturbed (HA-16)
sagt (Nc-3)	wolves (HC-2)
wurde ich (La-2)	even walk (after HA-3)

(See Figure 7, Part 1)

2. Characteristically, continuation is largely rising in German, largely falling in English. This appears first in the stressed syllable. The major portion of the stressed syllable of a continuation contour generally rises in German, falls in English.

Flucht (Ka-6)	home (HB-4)
gelebt (Kb-3)	perform (HA-7)
Gespräch (Ma-3)	debunkt (HA-13)
fragte (Lb-3)	certain (HA-2)

(See Figure 7, Part 2)

Furthermore, the perceived difference is greater than that shown by the frequency contour alone because the distribution of intensity bears on the rise in German, on the fall in English.

3. The rising versus falling characteristics of German and English also appear in the unstressed syllables of continuation that follow the stress. The post-stress syllables of a continuation contour are generally high in German, low in English.

werden (Ma-4)	later (HA-19)
wurde ich (La-2)	experiment (HA-8)
hängenblieben (Kb-4)	miserable (HA-15)

(See Figure 7, Part 3)

Note. When an English pattern of continuation is heard as rising, it is typically attributable, not to a rise on the stress at the start, but to a rising hook after the long fall, as in know (HA-1), certain (HA-2), experiments (HA-10), India (HB-3).

The bird shape used to symbolise the German contour could also be used for the English contour if the bird looked the other way, toward the left. Thus, the typical English pattern would begin with a high head, continue with a long falling neck and a low depression, sometimes followed by a rising tail, as in know (HA-1). Another possible symbol of the contour shapes could be the consonant S. Then the German contour would have a tilted S-shape and the English contour a tilted reversed-S shape.

4. In the expression of finality, the rise, in the overall falling patterns, is significant in German, not in English. Characteristically, in the German contour of finality, the head hump is preceded by a depression and the stress (the portion of highest intensity) bears on the rise (from depression to head). Characteristically, in the English contour, no depression precedes the head hump, the short rise (if any) is unstressed, and the highest intensity bears on the long fall.

sagt (Nc-3)	should (HC-1)
spricht (Nc-4)	wolves (HC-2)
Europas (Mc-2)	to work with (HC-4)
halluziniere ja nicht (Nc-1)	circumstances (HC-7)

(See Figure 8, Part 1)

5. In German finality, the unstressed syllables after the last stressed one characteristically break from the stress after a blank pitch interval. In English finality the corresponding unstressed syllables are, on the contrary,

in continuity with the falling curve of the stress -- they prolong this curve without a break.

Europas (Mc-2)	if needed (HC-3)
Hochdeutsch sprechen (Kc-1)	intensively (HC-5)
übernehmen wollte (Lc-1)	anthropology (HC-6)
verstanden zu werden (Kc-2)	human beings (HC-8)
Volkswagenwerke gekommen (Lc-2)	behavior at all (HC-9)

(See Figure 8, Part 2)

Note that the number of unstressed syllables of the above post-stress series is greater in German than in English. According to statistics, this is typical: for five minutes of recording each, subjects K, L, M, N average 2.6 syllables after the last stress; American subject H averages only 0.8 syllables. The German pattern of finality does not put the stress on the last possible stressed syllable of the sense group, as English normally does, but on the stressed syllable of the most important word among the last two or three. This is not entirely a word-emphasis process in German; it is largely a mechanical rhythmic habit.

6. The phonemic distinction of minor continuation versus major continuation is much clearer, and more often observed, in German than in English -- the acoustic correlates of German are more contrastive and more regularly present.

In German, major continuation has, on the average, a deeper depression of the back, a longer rise of the neck, a higher head, and a more rising beak of post-stress syllables.

Minor Continuation	Major Continuation
Flucht (Ka-6)	weiss (Nb-2)
verbannung (La-6)	entlassen (Lb-1)
Generaldirektor (Ma-9)	hängenblieben (Kb-4)

(See Figure 9, Part 1, left)

In English this distinction is generally ambiguous:

Minor Continuation	Major Continuation
know (HA-1)	home (HB-4)
certain (HA-2)	upright (HB-1)
experiments (HA-10)	India (HB-3)

(See Figure 9, Part 1, right)

7. The phonemic distinction of continuation versus finality is clearer in German than in English -- the acoustic correlates of German are more contrastive. In German, even though the rising portions of the patterns are alike for continuation and finality, there are two marked differences:

a) Continuation has no fall after the head; finality has one.

Continuation	Finality
weiss (Nb-2)	sagt (Nc-3)
gelebt (Kb-3)	Person (Nc-2)
geboren (Kb-2)	Europas (Mc-2)

b) The unstressed syllables which follow the last stress are high in continuation, low in finality.

Generaldirektor (Ma-9)	halluziniere ja nicht (Nc-1)
Volkswagenwerke (Ma-10)	verstanden zu werden (Kc-2)

(See Figure 9, Part 2, left)

Combining these two features, one can say that, in German, continuation is rising, finality is falling.

In English, both continuation and finality are falling, and even though finality falls lower than continuation, on the average, they are often hard to tell apart for a foreign ear,

home (HB-4)	should (HC-1)
debunkt (Ha-13)	wolves (HC-2)
upright (HB-1)	work with (HC-4)

unless there is a rising hook at the end of the continuation pattern, which is less than half of the time.

know (HA-1)	should (HC-1)
certain (HA-2)	needed (HC-3)
experiments (HA-10)	intensively (HC-5)

(See Figure 9, Part 2, right)

PEDAGOGICAL IMPLICATIONS

The American student of German must first acquire the phonetic habits a) of starting all stressed syllables with a depression and a stressed rise -- the drag effect will take care of itself if the rise is preceded by a depression -- and b) of divorcing the unstressed syllables from the stress in finality; then he must learn to distinguish c) continuation from finality by maintaining the post-stress syllables high for continuation and low for finality, and d) minor continuation from major continuation by producing a deeper depression, a higher rise in the stress, and more rise in the post-stress portion for major continuation. The combination of these four precepts roughly covers all the German characteristics of declarative intonation needing attention by an American student.

READING VERSUS SPONTANEOUS SPEECH

Three long, well-balanced sentences from Tonio Kröger, read by Thomas Mann himself in remarkable diction, are offered on Figure 4 to facilitate the comparison with the four improvisations by speakers K, L, M and N. The passage is on an LP record (Caedmon Tc1004). We note first that Mann's frequency range is much wider than that of the other speakers: 14 tones, more than 2 octaves, whereas the others varied from 6 tones (one octave) for K to 9.5 tones for M (1.5 octave). This would indicate a conscious effort of expression on the part of the speaker. He manifests all the intonation features found in improvisation (and no others), but they are generally more pronounced or better controlled. His depressions are deep before long rises. The bird shape is in nearly every stressed syllable. His distinctions between continuation and finality, or between minor and major continuation, are extremely clear. He has just over three a's per b (29 and 9), barely more than in improvised speech (36 and 13). The pitch break of the last syllables in finality occurs in each of the three sentence endings, but it is not so sharp as in the improvisations. He differs from the other speakers only by an organized distribution of word emphasis (6 e's), which is too well controlled to produce the sort of pitch disturbances found in the improvisations. In one case only, in JUpiter bart (e-3) does it result, as expected, in a subsequent fall of the unstressed syllable and a continuation beak which is lower than the head of the bird. In short, Thomas Mann provides abundant examples of all the declarative contours emerging from the present study, with only a suggestion of controlled emphasis and artificiality.

SUMMARY

To investigate the distinctive and non-distinctive characteristics of German declarative intonation, four passages of spontaneous speech were recorded, analyzed on spectrograms, and their frequency contours presented in graphic form (Figures 1 and 2). In addition, one sample of reading by Thomas Mann, and one of spontaneous speech by an American are offered for comparison (Figures 4 and 3).

A complete frequency contour of German continuation or finality can be divided into five significant segments, symbolized by a bird-shaped outline seen from profile. The five segments have the following pitch shapes: a tail fall, a back depression, a neck rise, a head hump, and a beak rise, or fall, or high plateau or low plateau. A statistical analysis of the occurrence, shape and function of each segment for the expression of minor continuity, major continuity, and finality has been made, and the perception of each segment in its function has been verified by synthesis on an artificial speech synthesizer of voback type. The main results are:

1. The frequency contours which phonetically characterize German declaration appear to be located a) in the back depression, by which most stressed syllables start (this depression before the neck rise seems to be responsible for the impression that stressed syllables are dragged in German); and b) in the low beak, or pitch break, of post-stress syllables in the finality contour.

2. The features of German frequency contours which meaningfully distinguish finality from continuation appear to be located mainly in the beak, the very end of the pattern, which is characteristically falling or low for finality, rising or high for continuation.

3. The German features which distinguish major continuation from minor continuation appear in all the segments of the bird shape. For major continuation, on the average, the back depression is deeper, the neck rise is longer, the head is higher, and the beak is higher and more rising. Note that, furthermore, the maximal intensity is on the rise of all German stressed syllables, even in the finality contour.

Finally, seven differences between German and American intonations of continuation and finality are described, with concrete examples from the graphic displays; and the controlled diction of Thomas Mann is favorably compared with that of spontaneous speech.

The composite contours of Figure 10 summarize our findings on German intonation for the expression of minor continuation, major continuation and finality, first without, then with unstressed syllables after the last stress of sense groups. They also show how English intonation would behave in realizing similar modes of expression under comparable conditions of stress and rhythm. In these composite shapes the location of maximal intensity (center of stress) is indicated by the thickness of the line.

Footnotes

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

[2] P. Delattre: "La leçon d'intonation de Simone de Beauvoir, étude d'intonation déclarative comparée," *The French Review*, Vol. 35, 1, pp. 59-67, October 1961.

[3] P. Delattre, C. Olsen, and E. Poenack: "A Comparative Study of Declarative Intonation in American English and Spanish," *Hispania*, Vol. 45, 2, pp. 233-241, May 1962.

[4] We wish to thank Professor E. Zwirner, Director of the Deutsches Spracharchiv, Münster, for providing us with recordings of spontaneous speech.

FIGURE CAPTIONS

SOME CHARACTERISTICS OF GERMAN INTONATION FOR THE EXPRESSION OF CONTINUATION AND FINALITY

Figure 1. Intonation graphs of spontaneous speech by two speakers of North German. Subject K, a woman, speaks very calmly. She has a range of 6 tones ($C_3$to C_4). Subject L, a man, speaks with emotion. He has a range of 8.5 tones (A_1 to D_3). a, b, c indicate the end of sense groups: a = minor continuation, b = major continuation, c = finality; e designates special emphasis.

Figure 2. Intonation graphs of spontaneous speech by two speakers of North German. Subject M, a man, speaks rationally. He has a range of 9.5 tones ($A^{\#}_1$ to F_3). Subject N, a man, speaks with intense emotion. He has a range of 10 tones (D_2 to $A^{\#}_3$), counting emotive word emphasis, and 7.5 tones (D_2 to F_3) not counting it. a = minor continuation, b = major continuation, c = finality, e = special emphasis.

Figure 3. Intonation graphs of spontaneous speech by a speaker of general American English. Subject H, a woman, speaks calmly and persuasively. She has a range of 6 tones (G_2 to G_3). A = minor continuation, B = major continuation, C = finality.

Figure 4. Intonation graphs of a recorded reading by Thomas Mann from Tonio Kröger. His diction is well controlled but exaggerated, with too many contrastive word emphases. In spite of the artificial effects due to reading, the same German pitch contours as in spontaneous speech appear clearly. Exaggerated diction shows in the frequency range, which is much wider than in spontaneous speech: 14 tones ($A^{\#}_1$ to D_4) counting the word emphases, 11 tones ($A^{\#}_1$ to $G^{\#}_3$) not counting them. a = minor continuation, b = major continuation, c = finality, e = special emphasis.

Figure 5. A complete presentation of the continuation contours produced by the four German speakers of Figures 1 and 2, divided according to the number of post-stress syllables. In the right hand column are the contours of word emphasis that are independent of a continuation pattern. Marked with e are the continuation patterns which are distorted by a word emphasis at the beginning. One can visualize the composite contour of each column.

Figure 6. A complete presentation of the finality contours produced by the four German speakers of Figures 1 and 2, divided according to the number of post-stress syllables. One can visualize the composite contour of each column.

Figure 7. A graphic comparison of German and English pitch shapes for stress and continuation. 1) All German stressed syllables characteristically start with a low depression and continue with a rise, even if a fall is to come later, as in grösste or sagt. English stressed syllables characteristically start with a high hump and continue with a fall, even if a rising hook is to come later, as in know. 2) The German stressed syllable of continuation characteristically rises after a low depression. The English stressed syllable of continuation characteristically falls after a high hump. 3) The German post-stress syllables of continuation characteristically remain high, forming a plateau or rise. The English post-stress syllables of continuation fall moderately, or hook up, as the last syllable of experiments.

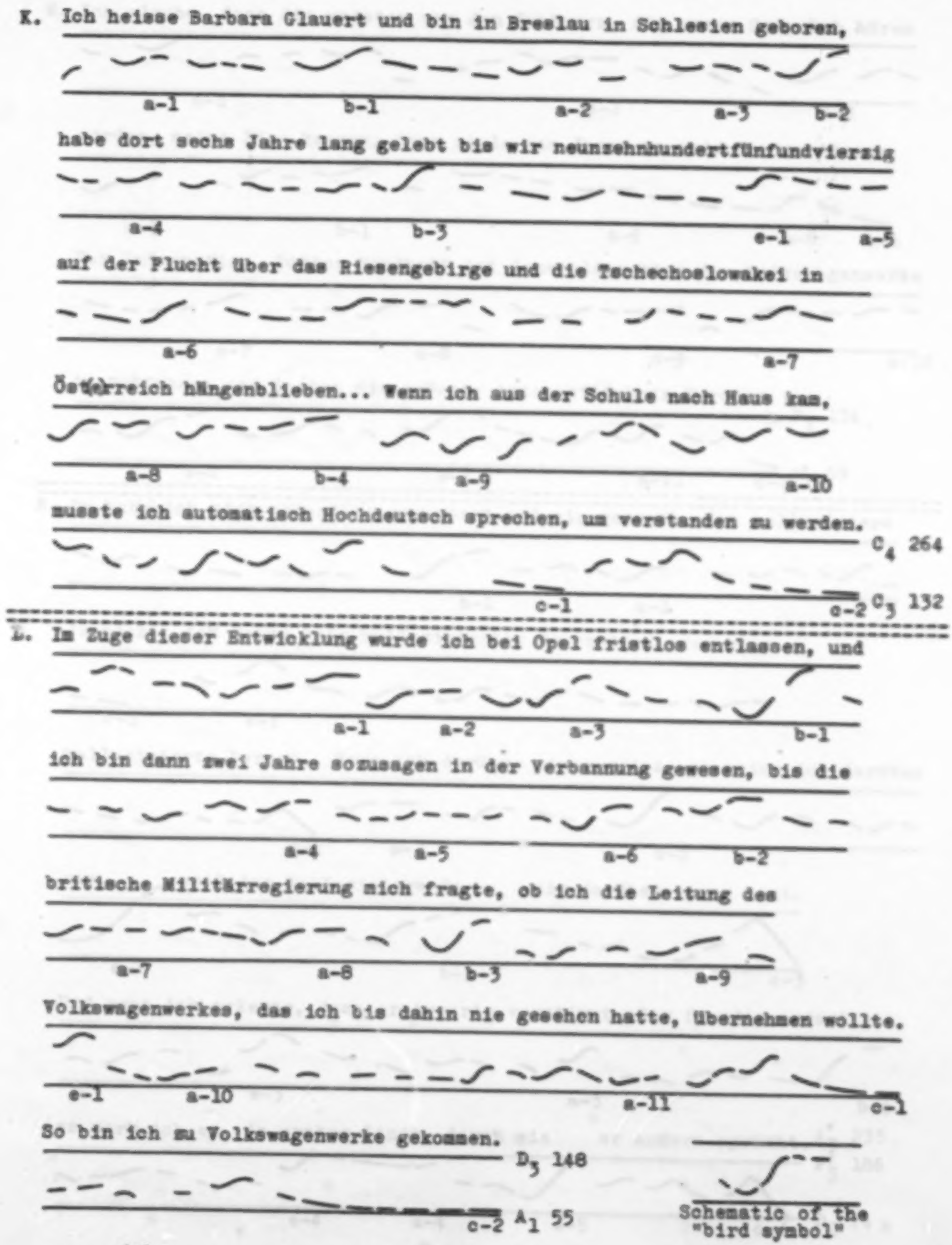
Figure 8. A graphic comparison of German and English pitch shapes for finality. 1) In German finality, the portion with maximal intensity is typically in the rise (for a stressed final syllable, as in sagt, therefore, the falling portion has low intensity). In English finality, the portion with maximal intensity is typically in the fall. 2) In German finality, the post-stress syllables are

typically separated from the stressed-syllable contour by a pitch break, which makes them sound as an independent semantic entity to a foreign ear. In English finality, the post-stress syllables are typically in continuity with the stressed syllable contour.

Figure 9. A graphic comparison of German and English pitch shapes for meaningful contrasts. 1) In German, major continuation is clearly distinguished from minor continuation by a deeper depression, a longer and higher rise, and more rise in the post-stress syllables. In English, this distinction is generally ambiguous; but it may be produced by a higher hump, a deeper fall, and/or a short rising hook after the fall. 2) In German, finality is clearly distinguished from continuation by a fall to complement the high stress or by a low plateau following the high stress. In English this distinction can be ambiguous, since both continuation and finality have falling contours; finality generally has a deeper fall and never ends with a rising hook as continuation may do.

Figure 10. A summary of the composite intonation contours of German which emerge from this study, and a comparison with the corresponding contours of English which emerge from previous studies, first without post-stress syllables (upper two lines), then with them (lower two lines). a, b, c indicate the key points of the pitch contours for the expression of minor continuation (a), major continuation (b), and finality (c) in German, and A, B, C the corresponding key points in English. Observe in both languages: the characteristic shape of all stressed syllables (S-shape in German, reversed-S shape in English); the pitch break of post-stress syllables in German finality and the continuity of the corresponding syllables in English; the characteristic shape of all continuations

(rising in German, falling in English); the contrast between continuation and finality (rise versus fall in German, different degrees of fall in English); the contrast between minor continuation and major continuation (more pronounced depression and rise in German, more pronounced fall and rising-hook in English when contrast is made); the location of maximal intensity (center of stress) in all syllables -- shown by the thickness of the line (on the rise, even for finality, in German, on the fall in English).

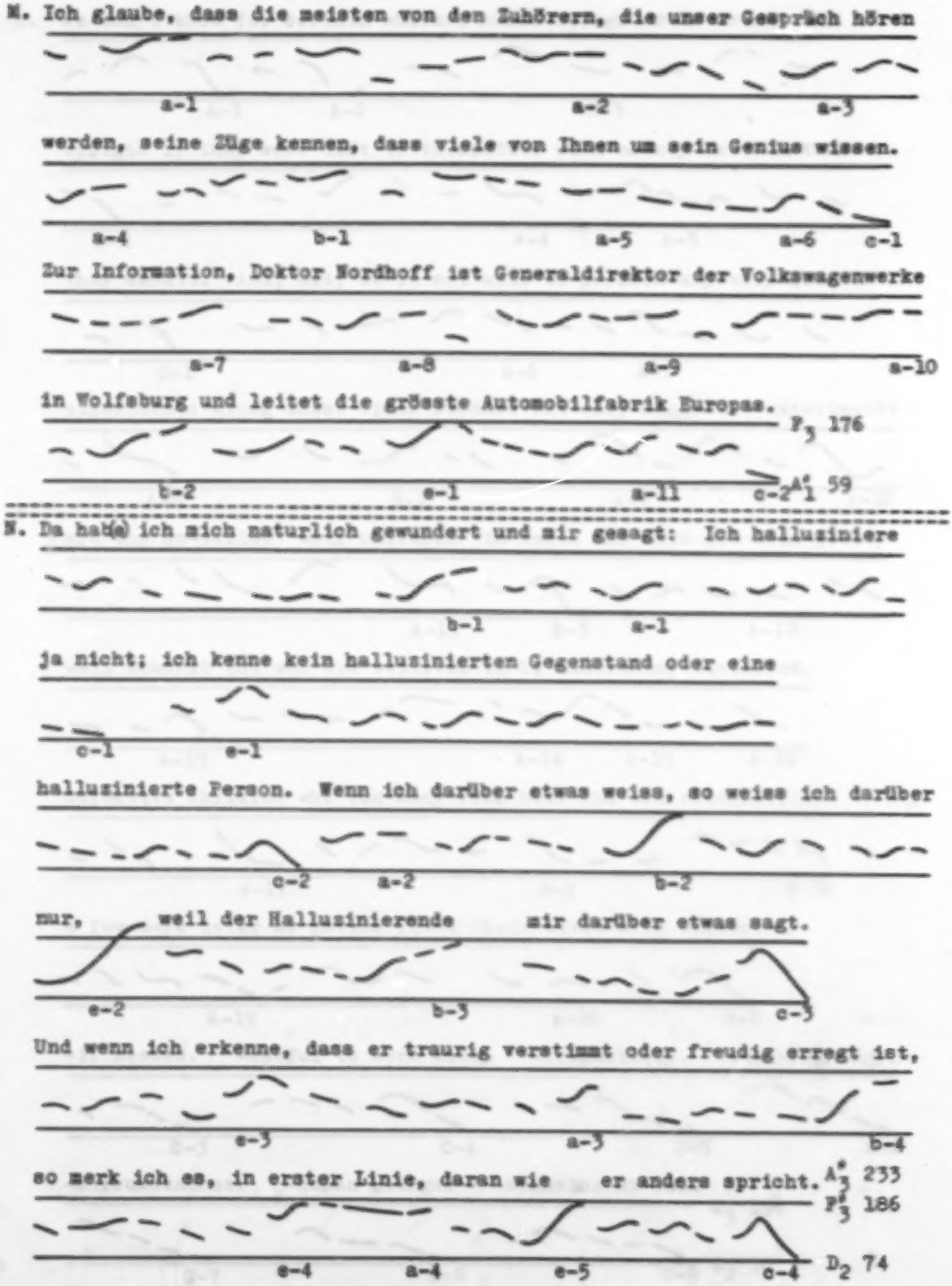


K. Ich heisse Barbara Glauert und bin in Breslau in Schlesien geboren, habe dort sechs Jahre lang gelebt bis wir neunzehnhundertfünfundvierzig auf der Flucht über das Riesengebirge und die Tschechoslowakei in Österreich hängenblieben... Wenn ich aus der Schule nach Haus kam, musste ich automatisch Hochdeutsch sprechen, um verstanden zu werden.

L. Im Zuge dieser Entwicklung wurde ich bei Opel fristlos entlassen, und ich bin dann zwei Jahre sozusagen in der Verbannung gewesen, bis die britische Militärregierung mich fragte, ob ich die Leitung des Volkswagenwerkes, das ich bis dahin nie gesehen hatte, übernehmen wollte. So bin ich zu Volkswagenwerke gekommen.

a-1
b-1
a-2
a-3
b-2
a-4
b-3
c-1
a-5
a-6
a-7
a-8
b-4
a-9
a-10
C_4 264
c-1
c-2
C_3 132
a-1
a-2
a-3
b-1
a-4
a-5
a-6
b-2
a-7
a-8
b-3
a-9
e-1
a-10
a-11
c-1
D_3 148
c-2
A_1 55
Schematic of the "bird symbol"

Figure 1

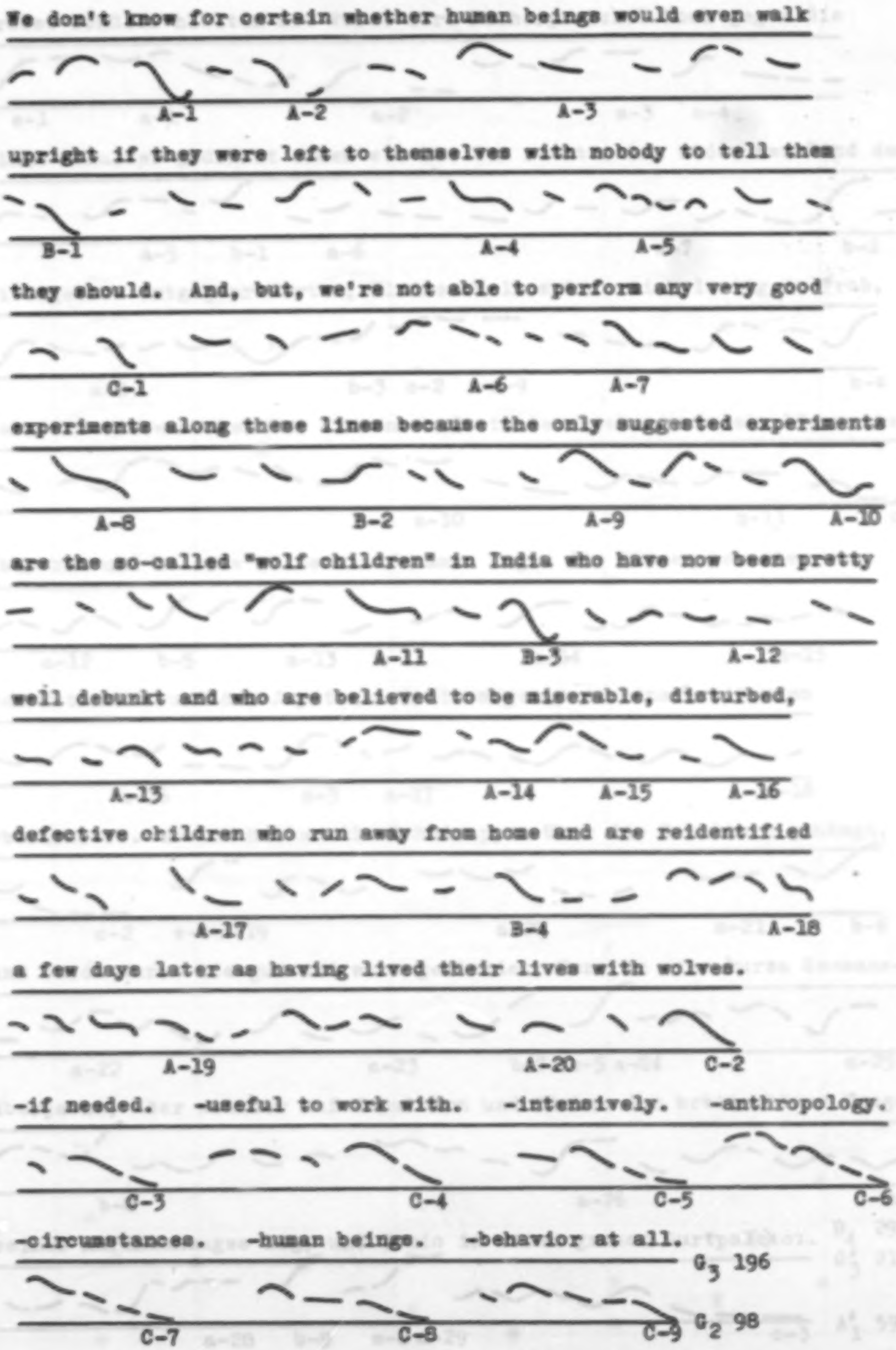


M. Ich glaube, dass die meisten von den Zuhörern, die unser Gespräch hören
werden, seine Züge kennen, dass viele von Ihnen um sein Genius wissen.
Zur Information, Doktor Nordhoff ist Generaldirektor der Volkswagenwerke
in Wolfsburg und leitet die grösste Automobilfabrik Europas.

N. Da hat(e) ich mich naturlich gewundert und mir gesagt: Ich halluziniere
ja nicht; ich kenne kein halluzinierten Gegenstand oder eine
halluzinierte Person. Wenn ich darüber etwas weiss, so weiss ich darüber
nur, weil der Halluzinierende mir darüber etwas sagt.
Und wenn ich erkenne, dass er traurig verstimmt oder freudig erregt ist,
so merk ich es, in erster Linie, daran wie er anders spricht.

a-1
a-2
a-3
a-4
b-1
a-5
a-6
c-1
a-7
a-8
a-9
a-10
F_3 176
b-2
e-1
a-11
c-2
A'_1 59
b-1
a-1
c-1
e-1
c-2
a-2
b-2
e-2
b-3
c-3
e-3
a-3
b-4
A'_3 233
F'_3 186
e-4
a-4
e-5
c-4
D_2 74

Figure 2



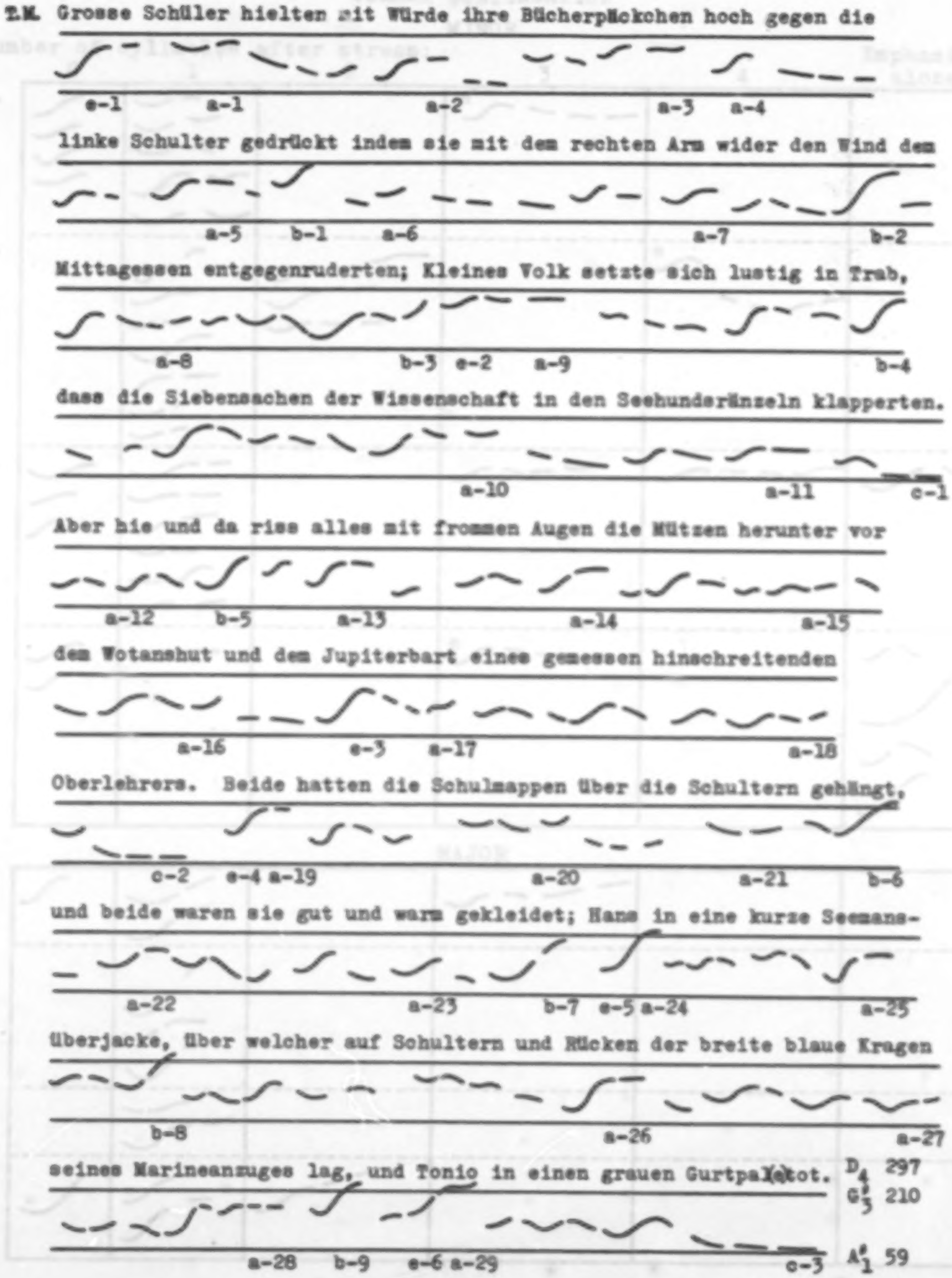
H. We don't know for certain whether human beings would even walk upright if they were left to themselves with nobody to tell them they should. And, but, we're not able to perform any very good experiments along these lines because the only suggested experiments are the so-called "wolf children" in India who have now been pretty well debunkt and who are believed to be miserable, disturbed, defective children who run away from home and are reidentified a few days later as having lived their lives with wolves.

-if needed. -useful to work with. -intensively. -anthropology.

-circumstances. -human beings. -behavior at all.

A-1
A-2
A-3
B-1
A-4
A-5
C-1
A-6
A-7
A-8
B-2
A-9
A-10
A-11
B-3
A-12
A-13
A-14
A-15
A-16
A-17
B-4
A-18
A-19
A-20
C-2
C-3
C-4
C-5
C-6
G_3 196
C-7
C-8
C-9
G_2 98

Figure 3



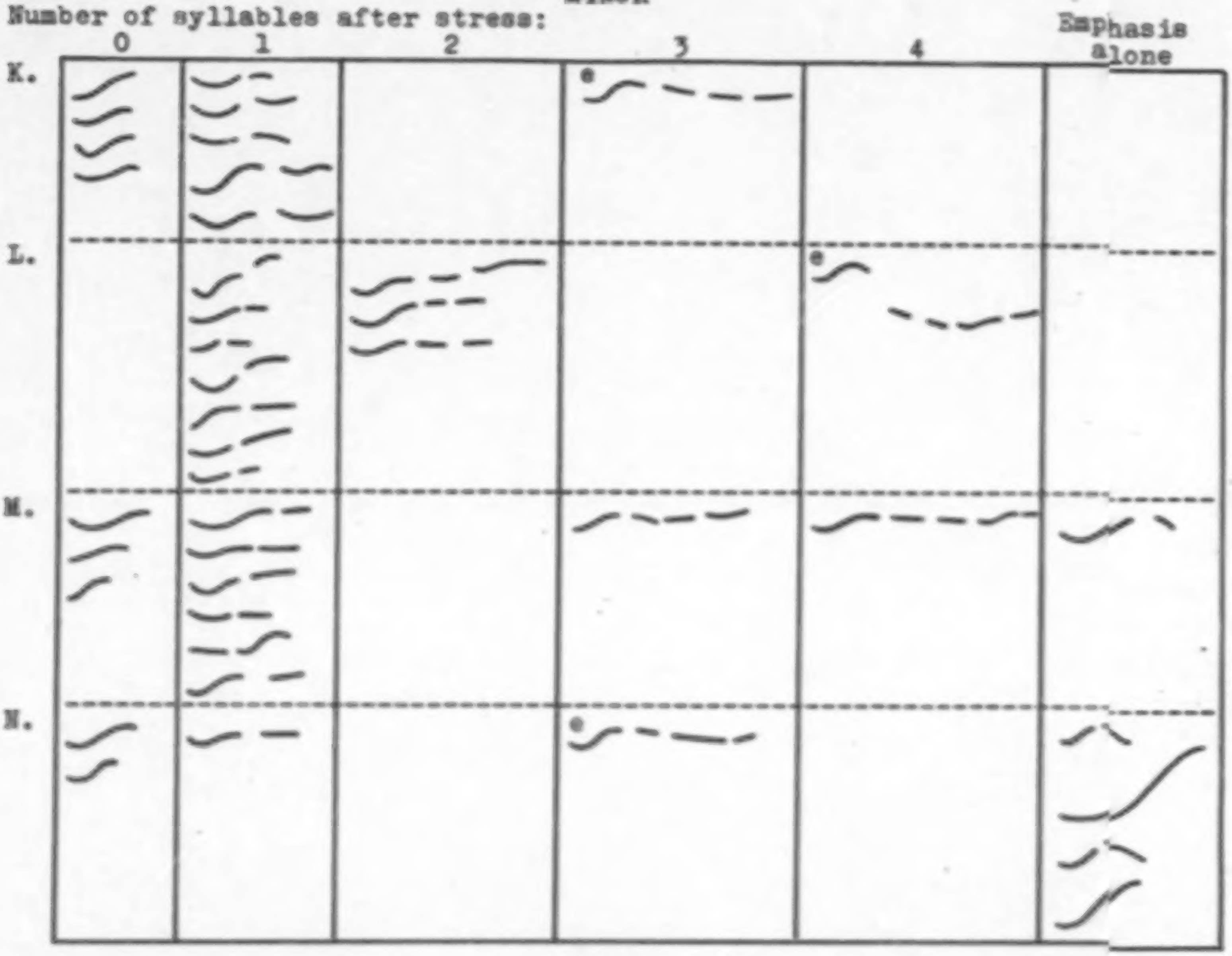
T.M. Grosse Schüler hielten mit Würde ihre Bücherpäckchen hoch gegen die
linke Schulter gedrückt indem sie mit dem rechten Arm wider den Wind dem
Mittagessen entgegenruderten; Kleines Volk setzte sich lustig in Trab,
dass die Siebensachen der Wissenschaft in den Seehunderänzeln klapperten.
Aber hie und da riss alles mit frommen Augen die Mützen herunter vor
dem Wotanshut und dem Jupiterbart eines gemessen hinschreitenden
Oberlehrers. Beide hatten die Schulmappen über die Schultern gehängt,
und beide waren sie gut und warm gekleidet; Hans in eine kurze Seemans-
überjacke, über welcher auf Schultern und Rücken der breite blaue Kragen
seines Marineanzuges lag, und Tonio in einen grauen Gurtpaletot.

e-1 a-1 a-2 a-3 a-4
a-5 b-1 a-6 a-7 b-2
a-8 b-3 e-2 a-9 b-4
a-10 a-11 c-1
a-12 b-5 a-13 a-14 a-15
a-16 e-3 a-17 a-18
c-2 e-4 a-19 a-20 a-21 b-6
a-22 a-23 b-7 e-5 a-24 a-25
b-8 a-26 a-27
a-28 b-9 e-6 a-29 c-3
D_4 297
$G^{\#}_3$ 210
$A^{\#}_1$ 59

Figure 4

GERMAN CONTINUATION

MINOR


Number of syllables after stress:
0
1
2
3
4
Emphasis alone
K.
L.
M.
N.
e
e
e

MAJOR

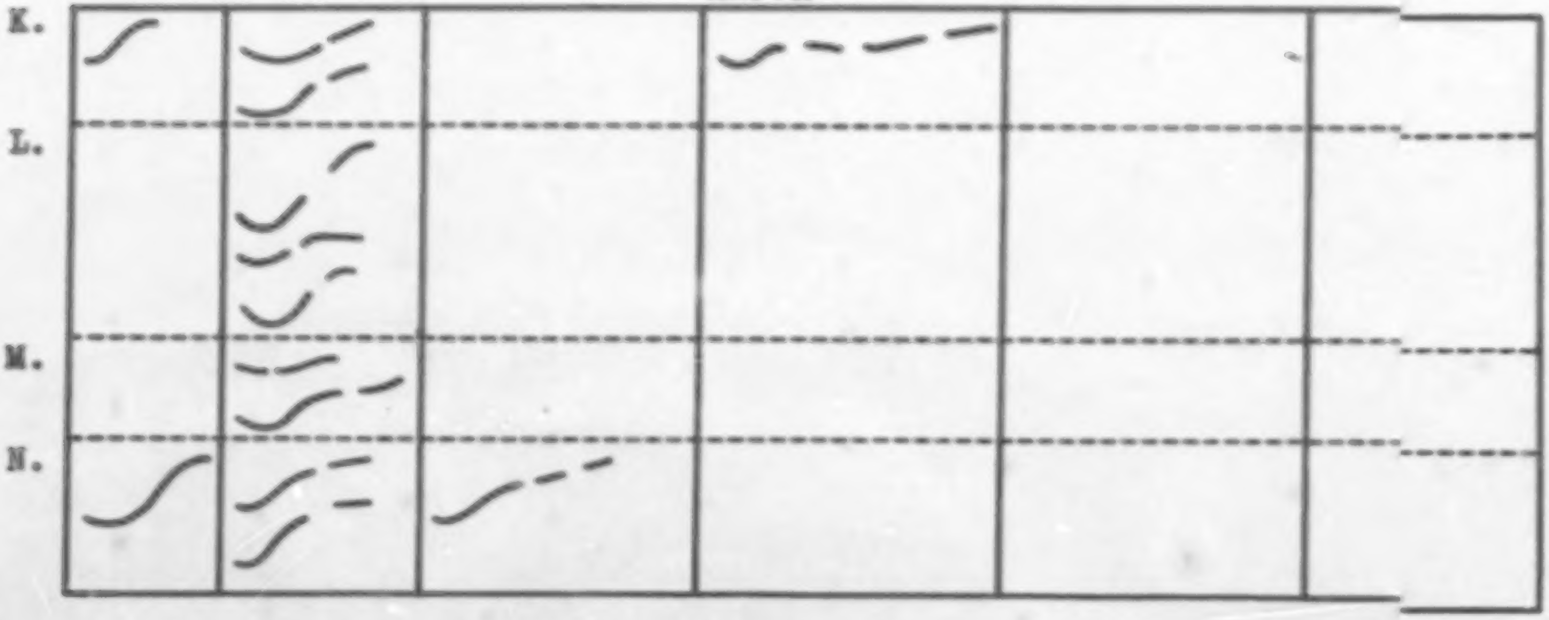
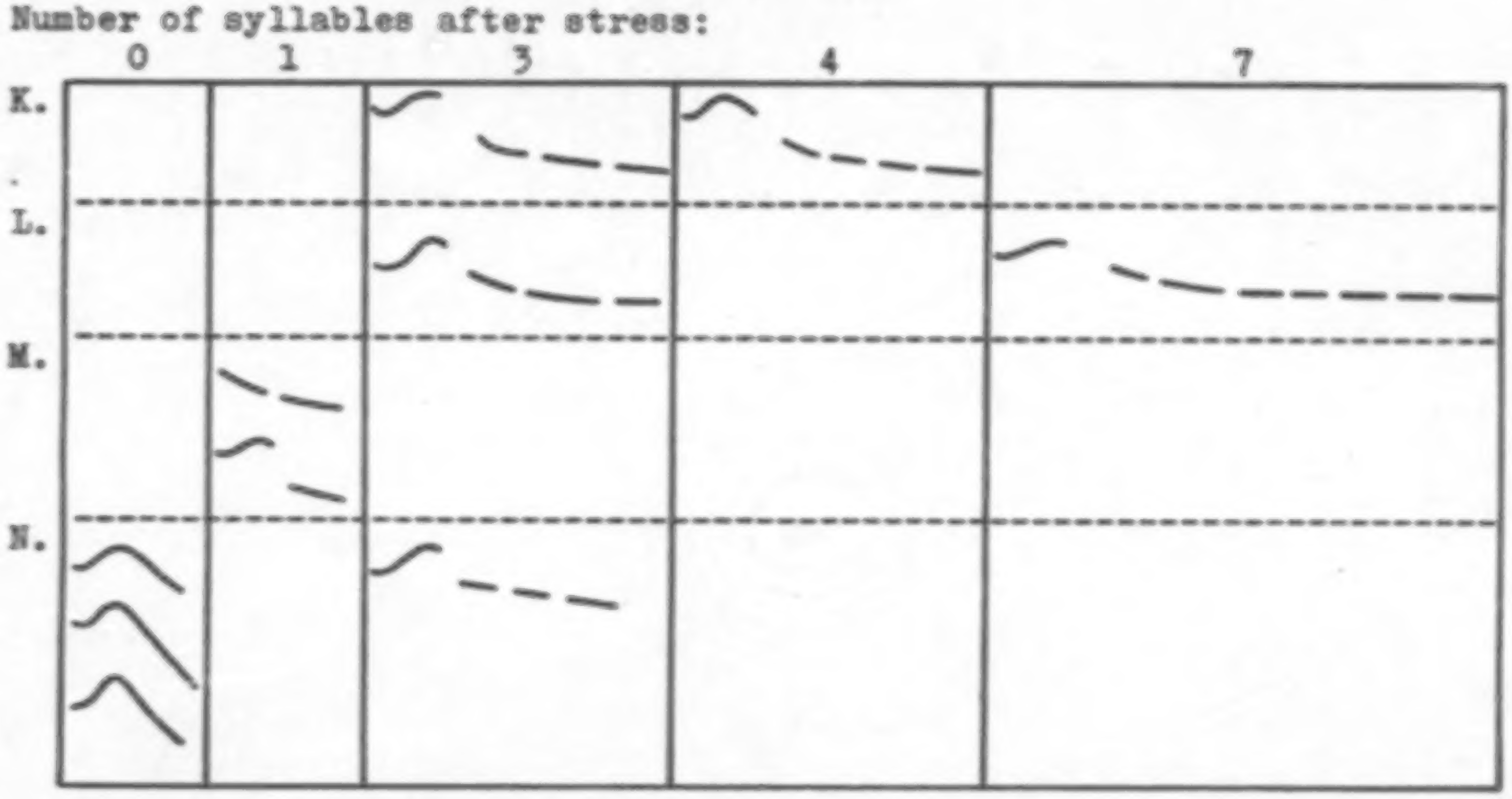


Figure 5


Number of syllables after stress:
0
1
3
4
7
K.
L.
M.
N.

GERMAN FINALITY

Figure 6

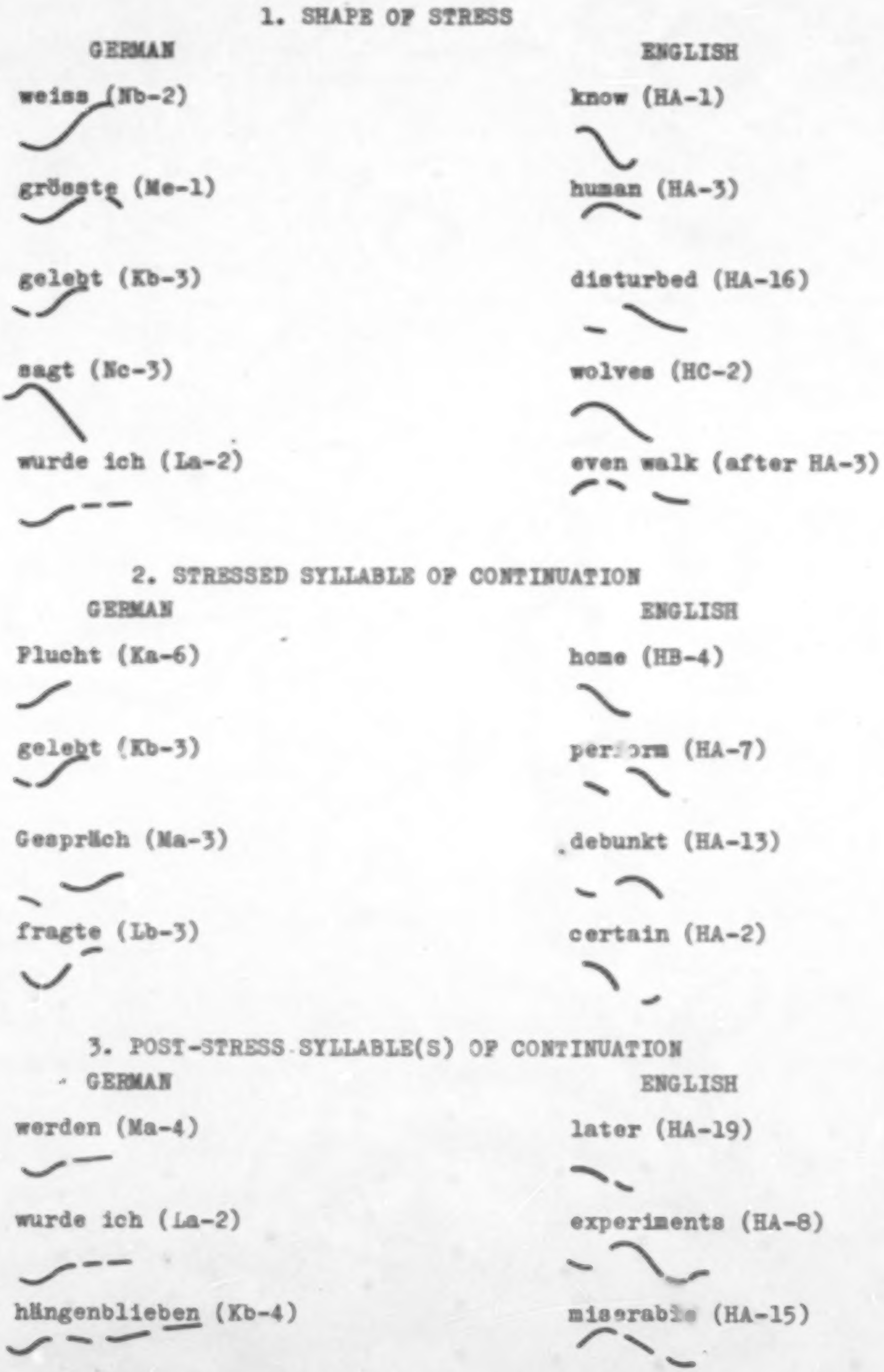
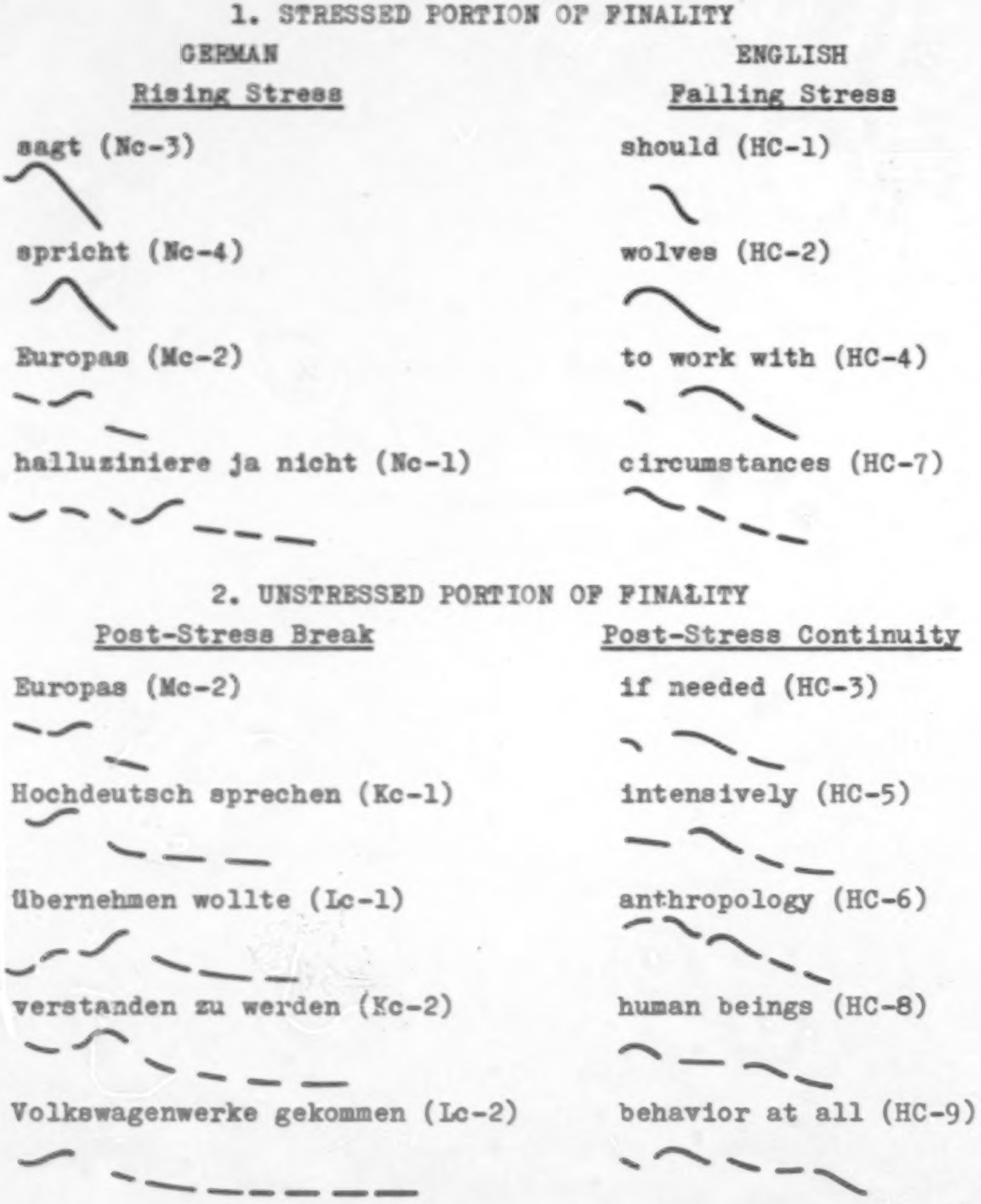

1. SHAPE OF STRESS
GERMAN
ENGLISH
weiss (Nb-2)
know (HA-1)
grösste (Me-1)
human (HA-3)
gelebt (Kb-3)
disturbed (HA-16)
sagt (Nc-3)
wolves (HC-2)
wurde ich (La-2)
even walk (after HA-3)
2. STRESSED SYLLABLE OF CONTINUATION
GERMAN
ENGLISH
Flucht (Ka-6)
home (HB-4)
gelebt (Kb-3)
perform (HA-7)
Gespräch (Ma-3)
debunkt (HA-13)
fragte (Lb-3)
certain (HA-2)
3. POST-STRESS SYLLABLE(S) OF CONTINUATION
GERMAN
ENGLISH
werden (Ma-4)
later (HA-19)
wurde ich (La-2)
experiments (HA-8)
hängenblieben (Kb-4)
miserable (HA-15)

Figure 7



1. STRESSED PORTION OF FINALITY

GERMAN Rising Stress	ENGLISH Falling Stress
sagt (Nc-3)	should (HC-1)
spricht (Nc-4)	wolves (HC-2)
Europas (Mc-2)	to work with (HC-4)
halluziniere ja nicht (Nc-1)	circumstances (HC-7)

2. UNSTRESSED PORTION OF FINALITY

Post-Stress Break	Post-Stress Continuity
Europas (Mc-2)	if needed (HC-3)
Hochdeutsch sprechen (Kc-1)	intensively (HC-5)
übernehmen wollte (Lc-1)	anthropology (HC-6)
verstanden zu werden (Kc-2)	human beings (HC-8)
Volkswagenwerke gekommen (Lc-2)	behavior at all (HC-9)

Figure 8

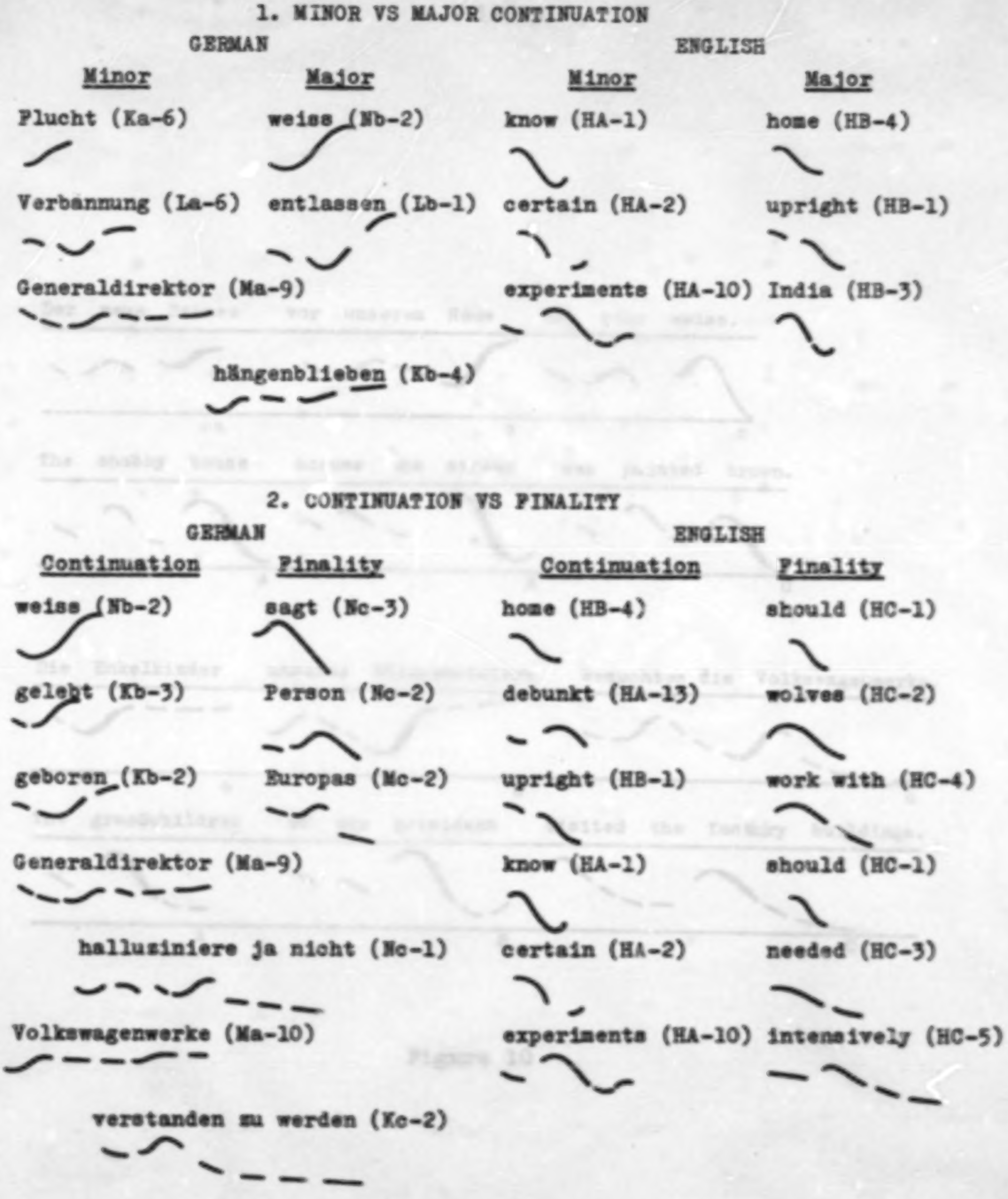

1. MINOR VS MAJOR CONTINUATION
GERMAN
ENGLISH
Minor
Major
Minor
Major
Flucht (Ka-6)
weiss (Nb-2)
know (HA-1)
home (HB-4)
Verbannung (La-6)
entlassen (Lb-1)
certain (HA-2)
upright (HB-1)
Generaldirektor (Ma-9)
experiments (HA-10)
India (HB-3)
hängenblieben (Kb-4)
2. CONTINUATION VS FINALITY
GERMAN
ENGLISH
Continuation
Finality
Continuation
Finality
weiss (Nb-2)
sagt (Nc-3)
home (HB-4)
should (HC-1)
gelebt (Kb-3)
Person (Nc-2)
debunkt (HA-13)
wolves (HC-2)
geboren (Kb-2)
Europas (Mc-2)
upright (HB-1)
work with (HC-4)
Generaldirektor (Ma-9)
know (HA-1)
should (HC-1)
halluziniere ja nicht (Nc-1)
certain (HA-2)
needed (HC-3)
Volkswagenwerke (Ma-10)
experiments (HA-10)
intensively (HC-5)
verstanden zu werden (Kc-2)

Figure 9

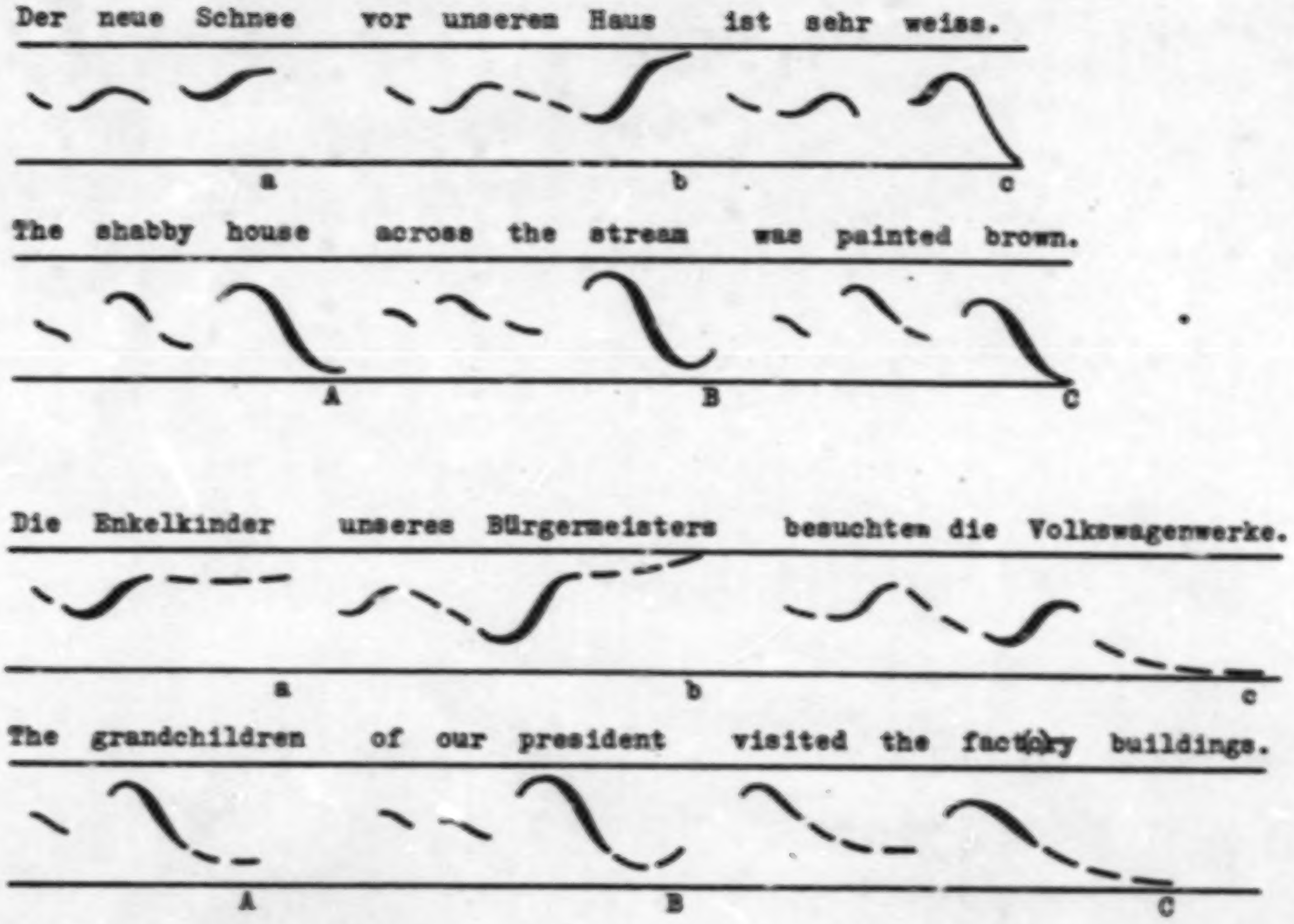

Der neue Schnee vor unserem Haus ist sehr weiss.
a
b
c
The shabby house across the stream was painted brown.
A
B
C
Die Enkelkinder unseres Bürgermeisters besuchten die Volkswagenwerke.
a
b
c
The grandchildren of our president visited the factory buildings.
A
B
C

Figure 10